Guidelines for designing and implementing a syndromic surveillance system

TRIPLE-S
Syndromic Surveillance Systems in Europe
# Table of contents

| Acknowledgments                              | iv  |
| Introduction                                 | 1   |
| 1.1 Why should I read these guidelines?      | 1   |
| 1.2 What is syndromic surveillance?          | 2   |
| 1.3 Why did syndromic surveillance develop?  | 3   |
| 1.4 What is the added value of Syndromic Surveillance? | 4   |
| 1.5 About these guidelines                   | 5   |
| 1.6 References                               | 8   |
| Setting up a syndromic surveillance system   | 11  |
| 2.1 Will syndromic surveillance meet my needs? | 11  |
| 2.2 Defining the purpose(s) of the syndromic surveillance system | 12  |
| 2.3 Identifying and contacting key partners  | 13  |
| 2.4 Check list for planning a syndromic surveillance system | 14  |
| 2.5 What resources are needed to create a syndromic surveillance system? | 15  |
| 2.6 References                               | 16  |
| Data collection, preparation and provision   | 17  |
| 3.1 Which data sources can I use?            | 17  |
| 3.2 Which variables should I collect?        | 39  |
| 3.3 Which coding system should be used?      | 48  |
| 3.4 How should data be recorded, transmitted and centralised? | 53  |
| 3.5 How should data be aggregated?           | 58  |
| 3.6 How can I ensure continuous access to data | 64  |
| 3.7 How can I ensure data confidentiality, security and quality? | 69  |
| 3.8 Minimum requirements for EU comparability | 73  |
| 3.9 Checklist                                | 74  |
| 3.10 Advice —“keep in mind”                  | 75  |
| 3.11 References                              | 75  |
## Data analysis

- General approach to analysing syndromic data
- Steps of data analysis
  - Preparation of data
  - Selecting and implementing appropriate statistical methods
  - From statistical signal to public health alert
  - Translation into public health action
  - Minimum requirements for EU comparability
  - Checklist for evaluation of data analysis
  - References

## Communication

- What is communication?
- What information for which stakeholders?
- Advice –“keep in mind”
- Minimum requirements for EU comparability
- Checklist for evaluation of communication
- References

## Evaluation

- What is evaluation?
- Why evaluate a syndromic surveillance system?
- What should be evaluated?
- How should the evaluation be done?
- How should the results be interpreted?
- Minimum requirements for EU comparability
- Checklist
- References

## Conclusion

## Glossary

## Appendix
The Triple-S project was coordinated by the French Institute for Public Health Surveillance (InVS).

Contributing authors

Context and introduction  Sylvia Medina (InVS, FR), Duncan Cooper (National Health Security, UK), Gillian Smith (Public Health England, UK), Jean-Baptiste Perrin, Céline Dupuy, Anne Bronner (French Agency for Food, Environmental and Occupational Health & Safety, FR), Hervé Chaudet (University of the Mediterranean Aix-Marseille II, FR).

Data collection, preparation and provision  Alex Elliot (Public Health England, UK), Celine Caserio-Schönemann (InVS, FR), Mikko Virtanen (National Institute for Health and Welfare, FI), Mark Kanieff (Istituto Superiore di Sanita, Italy), Céline Dupuy, Anne Bronner (French Agency for Food, Environmental and Occupational Health & Safety, FR).

Data analysis  Anne Fouillet (InVS, FR), Francoise Wuillaume (Institute of Public Health, BE), Arlene Reynolds (Health Protection Scotland, UK), Mikko Virtanen (National Institute for Health and Welfare, FI), Mark Kanieff (Istituto Superiore di Sanita, IT).

Dissemination and communication strategy  Marta Sala-Soler (InVS, FR), Anette Hulth (Swedish Institute for Communicable Disease Control, SE), Gillian Smith (Public Health England, UK), Céline Dupuy (French Agency for Food, Environmental Occupational Health & Safety, Anses, FR), Sylvia Medina (InVS, FR).

Evaluation  Luise Muller, Anne Mazick, Kåre Mølbak (Statens Serum Institut, DK), Alexandra Ziemann Thomas Krafft (Maastricht University, NL), Helena Medeiros (InVS, FR).

Glossary  Hervé Chaudet (University of the Mediterranean Aix-Marseille II, FR), Céline Caserio-Schönemann (InVS, FR), Céline Dupuy (French Agency for Food, Environmental Occupational Health & Safety, Anses, FR), Gillian Smith (Public Health England, UK).

Editorial team  Sylvia Medina, Marta Sala Soler, Helena Medeiros (InVS, FR), Mark Kanieff (Istituto Superiore di Sanita, IT).

We would like to thank the other members of the Triple-S consortium for their contribution to the project: Jim McMenamin (Health Protection Scotland, UK), Bart van der Zanden (Swedish Institute for Communicable Disease Control, SE), Suzanna Conti, Grazia Rago (Istituto Superiore di Sanita, IT), Satu Murtopuro (National Institute for Health and Welfare, FI), Anna Páldy, Janos Bobvos (National Institute of Environmental Health, HU), Liselotte van Asten (National Institute for Public Health and the Environment, NL), Paulo Lauriola (International Society of Doctors for the Environment, IT).

Triple-S would also like to thank the members of the advisory board: Howard Burkom (Johns Hopkins University for the International Society for Disease Surveillance – ISDS), Isabelle Devaux (European Centre for Disease Prevention and Control – ECDC), Fernanda Dorea (National Veterinary Institute, SE), Jurgita Kaminskaite (Executive Agency for Health and Consumers, Health Unit, European Commission), Franz Karcher (DG Health and Consumers, European Commission), Bettina Menne (World Health Organization Regional Office for Europe) and other contributors who helped enrich the guidelines including Loic Josseran (Université de Versailles-Saint-Quentin, FR), Michael Saklad (Saklad Consultants for communications strategy, FR), Anne-Catherine Viso (InVS, FR). We would also thank Sarah Moncrieff and Anette Hulth for formatting and designing this document.
Introduction

1.1 Why should I read these guidelines?

Have any of the following scenarios occurred or could occur in your country or area? And if so, what information would be available to assess the impact of the described event?

- A large fire at a chemical plant releases dangerous gases and chemicals into the atmosphere. The authorities, the media and the public need to know whether the chemical release is causing health problems in the community. How does the local public health department assess the immediate health impact of the fire or confirm that there has been no impact?

- Anecdotal reports suggest that there has been a sharp increase in the occurrence of diarrhoea in the community. To confirm these reports, the authorities wait for biological samples to be collected and for the results of laboratory testing. But what is the immediate impact on the population in terms of the numbers, age groups and areas most affected?

- There is an unexpected heat wave across the country. How can you ensure that widespread heat stroke, heat-related illness and death are not occurring?

- There is an important mass gathering event in your country, is the traditional surveillance system sufficient to monitor in real-time the potential health impacts during the event?

- Several farmers and veterinarians report abnormal calf mortality. How can this increase be confirmed and how can it be determined where and when it started?

- A new disease appears in your country. How can you ensure that you will be able to detect additional cases early enough or at all?

If the answer is “yes”, you should read these guidelines. They present different examples where syndromic surveillance can be of value in supporting existing surveillance systems and public health monitoring.

These guidelines describe what syndromic surveillance is and its added value for public health surveillance. They provide a practical description of how to set up a system, which data sources to use, how to analyse the data, report the findings and evaluate the system’s usefulness.
The Triple-S project defines syndromic surveillance as follows:

“The real-time (or near real-time) collection, analysis, interpretation and dissemination of health-related data to enable the early identification of the impact (or absence of impact) of potential human or veterinary public health threats which require effective public health action.

Syndromic surveillance is not based on laboratoryConfirmed diagnoses of disease but on non-specific clinical signs, symptoms and proxy measures for health (for example, absenteeism, drug sales, animal production collapse) that constitute a provisional diagnosis (or ‘syndrome’).

The data are usually collected for purposes other than surveillance and, where possible, are automatically generated so as not to impose an additional burden on data providers. Although syndromic surveillance tends to be non-specific it can be sensitive and rapid, and can augment and complement the information provided by laboratory-based surveillance systems”. (Triple S project, 2011)

In human health, syndromic surveillance is characterised by the timeliness of data collection and analysis, and the type of indicators that are monitored, i.e. unspecific health indicators or proxies.

In animal health, the definition of syndromic surveillance is less restrictive regarding the timeliness and automation of data collection and analysis. Timeliness and automation are perceived more as objectives (often not attained), rather than as inherent characteristics.

Syndromic surveillance covers a wide range of purposes, in particular:

- **Detection**: to detect an unknown, unexpected or emerging public health threat
- **Early warning**: to detect the start of an expected event (for example, a seasonal event such as influenza)
- **Situational awareness**: to quantify and monitor the impact of an identified public health threat; and demonstrate the lack of public health impact of a known threat (provide reassurance).

Syndromic surveillance is part of public health surveillance1, the latter being an umbrella term that also includes ‘specific surveillance’ of identified human and animal diseases, disease-causing agents and health risks. Specific surveillance generally uses laboratory reports, notifications of diagnoses by clinicians and case reporting. Syndromic surveillance is intended to complement specific surveillance. It is a supplementary tool that can detect changes or events that would either not have been detected or would have been detected later by a specific surveillance system.

In these guidelines, when we refer to public health surveillance we include both human and veterinary public health surveillance.

1. Public health surveillance, as defined by the World Health Organization (WHO, 2012), is the continuous and systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice. Such surveillance can: serve as a warning system for impending public health emergencies; document the impact of an intervention, or track its progress towards specified goals; and monitor and clarify the epidemiology of health problems to allow priorities to be set and to inform public health policy and strategies.
In recent years, syndromic surveillance systems that collect, analyse and interpret pre-diagnostic data have been developed to estimate the health status of the population in a timely way. ‘Pre-diagnostic’ means before a confirmed diagnosis is made, or the information that has been collected if a diagnosis is never made. In veterinary medicine formal diagnosis is probably less often reached than in human medicine, and laboratory tests are generally used only if there is a good cost-benefit ratio, especially for livestock. Syndromic surveillance using pre-diagnostic data seems thus particularly pertinent for animal health.

The impetus for the development of syndromic surveillance was the need to measure population health rapidly in response to the emergence of diverse public health threats such as bioterrorism and SARS. Much of the initial enthusiasm and support for syndromic surveillance was based on the need to detect such threats. In the United States particularly, many county-, city- or state-wide syndromic surveillance systems were established in the early 2000s. It also proved useful in the absence of laboratory capacities for case confirmation.

In addition to this need was the parallel development of health care systems that collected data automatically as part of routine clinical practice. There was, therefore, the potential to automatically obtain health care attendance data on clinical symptoms and morbidity. These data could be aggregated at a population level and analysed swiftly in order to provide a timely ‘snapshot’ of the health of the population.

In animal health, crises such as the classical swine fever (the Netherlands, 1997–1998), foot-and-mouth disease (Great Britain, 2001) and, more recently, the BTV-8 (first detected in the Netherlands, 2006) and Schmallenberg virus emergencies (first detected in Germany, 2011), that spread across Europe, have demonstrated the difficulty in detecting emerging diseases presenting with non-specific clinical symptoms. The importance of detecting these occurrences in a timely manner lies in their impact on animal populations and the potentially severe economic losses. For example, the losses due to the Netherlands’ outbreak of classical swine fever were estimated in 1997 at US $2.3 billion (Meuwissen et al., 1999).

In light of these circumstances, syndromic surveillance has been growing steadily in Europe over the past years (Conti S et al., 2013; Dupuy C et al., 2013). Today, over 15 European countries conduct syndromic surveillance on a local, regional or national level. They draw on diverse data sources and are aimed at a variety of health risks: from SARS to H1N1, from heat waves to volcanic ash clouds, from threats of bioterrorist attacks at G8 summits to mass gatherings at Olympic Games, and from financial crises to refugee camps. Syndromic surveillance systems range from simple (for example, based on a single syndromic indicator in one area) to complex (multiple syndromic indicators covering an entire country).
A major contribution of syndromic surveillance is that it usually employs ‘real-time’ or ‘near real-time’ data analysis with the intention of providing a more timely warning of infectious disease outbreaks, epidemics or pandemics, potential bio-terrorist attacks, environmental events or major chemical or industrial incidents.

One of the most useful benefits of syndromic surveillance on a day-to-day basis is situational awareness. This can aid public health institutions monitor key aspects of a threat, including its severity and geographic extent, the population at risk and the dynamics of its evolution. It can also identify increases in illness rates (or proxy indicators of illnesses) before diagnoses are confirmed and reported to public health agencies.

Some published examples of how syndromic surveillance has been used to benefit public health response are provided in Table 1.1.

The benefits, outlined in Box 1.1, can be used for system evaluation, described in more detail in Section 6 on evaluation.

### Olympic Lessons

Syndromic surveillance can provide additional surveillance support during a major national mass gathering, such as the Olympic and Paralympic Games, as part of a wider enhanced public health surveillance programme. Such surveillance systems can provide additional reassurance that nothing is happening with communities in close proximity to the mass gathering and can also contribute additional surveillance intelligence in the event of an outbreak or incident.
BOX 1.1 Main benefits of syndromic surveillance systems for public health surveillance

Based on non-specific and wide-ranging health indicators, and can so help detect a broad range of events (expected or unexpected).

Often based on symptoms reported to health services in the previous 24 hours, and may thus indicate an increase in the occurrence of symptoms before laboratory confirmation.

Can detect symptoms of known or emerging diseases for which there are no existing disease-specific surveillance systems.

Can validate and support alerts generated by other surveillance systems across Europe (for example, event- and indicator-based surveillance systems).

Can be used for short-term surveillance during mass gatherings (for example, sporting events).

Can identify suspected infected persons for microbiological sampling (for example, during influenza epidemics).

Can be used for surveillance of earthquakes, tsunamis and other catastrophic events that can affect and sometimes displace large populations.

Can provide (near) real-time information for media enquiries about population health in the event of a public health emergency.

Can identify potential signals and provide outputs that may help decision makers take early action.

Can strengthen public health networks by following up syndromic surveillance alerts.

Can, in some cases, be relatively inexpensive to create (compared to disease-specific surveillance systems), because syndromic surveillance is often based on data that is already collected.

In what context were these guidelines developed?

Over the last decade an informal network of European public health professionals interested in syndromic surveillance has developed syndromic surveillance systems in their countries. Based on exchanges among these professionals and on questions from their governments about syndromic surveillance in other European countries, this network answered a call from the Executive Agency for Health and Consumers (EAHC) to answer the questions set out in the sections above.

Who developed these guidelines?

The Triple-S project (Syndromic Surveillance Systems in Europe) was launched in 2010 for a three-year period. Coordinated by the French Institute for Public Health Surveillance (InVS), it involved 24 organisations from 13 countries.

The objectives of the Triple-S project were to:

- Review and analyse syndromic surveillance activities across Europe and create an inventory of such activities
Facilitate networking and knowledge exchange among those parties active and/or interested in syndromic surveillance

Develop guidelines to support the implementation of syndromic surveillance systems in European Union member states, whilst respecting the diversity of health systems and data sources in these countries

Propose a European strategy for syndromic surveillance.

The partners involved in the Triple-S project worked together to develop these guidelines, which can be read in conjunction with other outputs of the project (such as fact sheets and separate guidelines on the assessment of data sources).²

Who are these guidelines for?

These guidelines are intended for public health professionals and epidemiologists who use human or animal health surveillance as part of their work and want to develop or improve syndromic surveillance systems.

What are the purpose, scope and limitations of these guidelines?

These guidelines provide evidence-based practical recommendations and suggestions for designing, implementing and improving syndromic surveillance systems. They include examples of past and on-going initiatives from several European countries. They draw on the experiences of those working in many different types of systems at varying stages of development. While the guidelines are based on European experiences, their main principles are valid globally.

The guidelines describe minimum requirements for developing syndromic surveillance systems and reporting surveillance findings. They also provide additional suggestions for improving systems beyond what is defined as the minimal level. They aim to encourage a common understanding of the structure and utility of syndromic surveillance systems across Europe, and to improve communication between European member states on critical public health threats. Because member states have differing health care infrastructures and varying objectives, resources, etc., these guidelines do not propose a single surveillance model for Europe or even for individual countries.

How are these guidelines organised?

The way in which these guidelines are organised follows the order of the steps necessary for designing and implementing a syndromic surveillance system (Figure 1.1):

- Getting organised:
  - determining whether syndromic surveillance meets your needs
  - defining the purpose(s) of the system
  - identifying and contacting key partners
- Collecting and preparing the data

² Fact sheets and guidelines for the assessment of data sources can be found on the Triple-S Website: www.syndromicsurveillance.eu.
• Analysing the data
• Disseminating the findings to support public health action
• Evaluating the system.

Please note that it is essential to revisit these steps regularly in order to re-evaluate the system once it is in place.

To help the reader, each section is illustrated with local or national experiences from Triple-S partners in animal or human syndromic surveillance and ends with ‘minimal requirements for EU comparability’ and a ‘checklist’. There are also some examples throughout each chapter referring to specific requirements of mass gathering of data, drawing on the experience of Triple-S partners involved in coordinating a syndromic surveillance programme for the London 2012 Olympic and Paralympic Games.

'Human' syndromic surveillance systems
Examples are shown in blue boxes.

'Animal' syndromic surveillance systems
Examples are shown in green boxes.
List of related documents produced by Triple-S

Summary sheets: provide a brief description of the purposes and data sources of existing systems in Europe.

Fact sheets: a series providing summaries of each document produced by Triple-S, including the guidelines.

Guidelines for assessing data sources: these describe all types of syndromic data sources. They are based on evidence from the published literature and from site visits made during the Triple-S project to learn about syndromic surveillance systems developed in a number of member states (Ziemann A, 2013).

Glossary: at the end of this document contains the main terms used in these guidelines.


Before deciding whether or not to develop a syndromic surveillance system, you need to ask yourself if it will actually help with the problem(s) you are trying to address.

To answer this, the following questions should be considered:

- Are there gaps in the existing surveillance systems that prevent events from being detected early enough (or at all) for a public health response to be effective?
- Can existing surveillance systems help address these gaps?
- Can syndromic surveillance address these gaps better or more rapidly than traditional surveillance?
- Can a syndromic surveillance system be implemented using the existing health care infrastructure, resources and data sources?

Two frameworks for considering the usefulness of syndromic surveillance are presented in the figures below. In particular, the types of events for which syndromic surveillance may be useful (Figure 2.1) and its timeliness compared to other systems (Figure 2.2).

**Figure 2.1 Decision tree for determining if syndromic surveillance meets your needs**
When designing and implementing a syndromic surveillance system in your country, it is essential to define clearly the purpose(s) of the system (see ‘Introduction’). Most syndromic surveillance systems serve multiple purposes. The added value and feasibility of European comparability may also be considered at this stage. Although a system may be set up initially for a single purpose (for example, to monitor health during a mass gathering event), it can be expanded to serve other purposes.

In defining the purpose of the system it is worth considering whether there is a link or synergy with existing traditional system(s) (for example, established influenza surveillance systems) and if this synergy would be useful in achieving the system’s aims (Dupuy C, 2012). The same applies for a synergy between animal and human surveillance systems. The benefits of synergy include: pooled expertise and resources; cross-validation of alerts from other surveillance systems; ensuring a continuous surveillance in case of
disrupted systems or overload due to an increase in the number of patients during a crisis; and the ability to describe events from a syndromic to a more specific level.

It is also necessary to consider the nature of the public health event(s) that the system is intended to monitor:

- A symptom or clinical syndrome (for example, vomiting, gastroenteritis or bloody diarrhoea, cough, fever, flu-like illness)
- A proxy indicator for a specific disease (for example, absenteeism for a respiratory disease, sale of over-the-counter flu remedies, or percentage of bovine abortion for Q fever infection)
- A disease or a group of specific diseases (for example, bronchitis, upper and lower respiratory tract infections, respiratory diseases)
- An emerging or unknown human, animal or zoonotic disease (for example, a novel respiratory illness, pandemic influenza, Schmallenberg virus)
- The impact of an environmental event (for example, heat wave-related effects, allergy or asthma associated with pollens or air pollution)
- A deliberate or accidental release (for example, chemical, biological, radiological or nuclear)
- The adverse effects of industrial (chemical, nuclear) accidents.

**Olympic Lessons**

When establishing a syndromic surveillance system specifically for a mass gathering it is important to define its purpose with respect to the structure of the event, and also to bear in mind the requirements for the ‘legacy syndromic surveillance system’ post-mass gathering i.e. what will the syndromic surveillance need for routine surveillance of seasonal pathogens?

---

2.3 Identifying and contacting key partners

In the early stages of designing a syndromic surveillance system, a steering group should be created. This should include data providers, data managers, statisticians, epidemiologists, funders/sponsors and public health staff as they are going to play a role in the system.
## 2.4 Check list for planning a syndromic surveillance system

Although the following sections in these guidelines will take you through specific issues regarding how to implement a syndromic surveillance system, the list below provides a non-extensive list of ‘tips’ for planning and implementation.

<table>
<thead>
<tr>
<th>OPERATIONAL</th>
<th>Identify the health care infrastructure(s) on which the syndromic surveillance system could rely, including human, financial and other resources.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ensure that the data providers are willing to participate and that providing data does not constitute an additional burden to their work.</td>
</tr>
<tr>
<td></td>
<td>Identify key partners who will support the system (internal and external stakeholders).</td>
</tr>
<tr>
<td></td>
<td>Clearly define the roles, responsibilities and expectations of internal and external stakeholders.</td>
</tr>
<tr>
<td></td>
<td>Create a management structure for the system, including a steering committee (or equivalent) and governance issues.</td>
</tr>
<tr>
<td></td>
<td>Identify links with other public health surveillance systems.</td>
</tr>
<tr>
<td></td>
<td>Identify links between human health and animal health surveillance systems.</td>
</tr>
<tr>
<td></td>
<td>Define a protocol for investigating alarms and issuing alerts.</td>
</tr>
<tr>
<td></td>
<td>Be flexible and take advantage of opportunities that arise (for example, a large mass gathering may add impetus to the development of a system).</td>
</tr>
<tr>
<td></td>
<td>Understand your syndromic surveillance system outputs before you start alerting.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LEGAL CONTEXT AND GOVERNANCE</th>
<th>Identify barriers that may hinder implementation (for example, conflict of interests, scarce resources).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identify key aspects of data governance, in particular, legal issues and data confidentiality.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENERAL</th>
<th>Think big, start small, and get key people on your side.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Set realistic expectations of what the system will and will not deliver.</td>
</tr>
<tr>
<td></td>
<td>Aim to provide a service not a system.</td>
</tr>
<tr>
<td></td>
<td>Understand the health care system, the use of health care by the public and how it changes over time (for example, during a pandemic).</td>
</tr>
<tr>
<td></td>
<td>Define the criteria for evaluating the success of the system and who will evaluate it.</td>
</tr>
</tbody>
</table>
It is difficult to estimate accurately the human and financial resources required to set up a syndromic surveillance system, since these are influenced by:

- the size of the system (for example, national or regional)
- the complexity of the data (for example, is mapping or translation of the coding required?)
- centralisation of the data transmission (for example, are the data submitted from one centralised data provider or from multiple health care providers?)
- the infrastructure of the ‘host’ organisation responsible for the syndromic surveillance system (for example, resources may already be available to support such initiatives).

Table 2.1 provides an estimate of the resources needed to set up a system that meets the minimum requirements of a syndromic surveillance system and a more complex enhanced system. This information provides basic guidance; requirements will differ depending on the existing infrastructure of the host organisation.

The system should have the resources needed to conduct internal and external evaluations (Table 2.2)

<table>
<thead>
<tr>
<th>Level</th>
<th>Human resources</th>
<th>Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum level</td>
<td>Data analysts (2 persons)</td>
<td>Computers</td>
</tr>
<tr>
<td></td>
<td>Scientific assistant (1 person)</td>
<td>Software</td>
</tr>
<tr>
<td></td>
<td>Senior epidemiologist* (1 person)</td>
<td>Phones</td>
</tr>
<tr>
<td></td>
<td>IT (1 person)</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>Communication**</td>
<td></td>
</tr>
<tr>
<td>Enhanced level: additional suggested resources to improve the system or manage multiple systems</td>
<td>Medical epidemiologist (1 person)</td>
<td>Computers</td>
</tr>
<tr>
<td></td>
<td>Scientific assistant (2 persons)</td>
<td>Servers</td>
</tr>
<tr>
<td></td>
<td>Experienced statistician (1 person)</td>
<td>Offsite backup servers</td>
</tr>
<tr>
<td></td>
<td>Administrative support (1 person)</td>
<td>Software</td>
</tr>
<tr>
<td></td>
<td>IT (1 person)</td>
<td>Phones</td>
</tr>
<tr>
<td></td>
<td>Business and communication support (1 person)**</td>
<td>Other</td>
</tr>
</tbody>
</table>

* The epidemiologist should have substantial experience in statistics and data analysis.
** Communication should form part of the system’s work.
*** Ideally, there should be a clear link between the organisation’s communications or media professional and the system owners in both day-to-day communication and in time of crisis/incident. The communication strategy of the syndromic surveillance system should be linked to that of other existing surveillance systems (sharing media professionals if possible). If a dedicated website is planned, sufficient resources are needed for updating and maintaining the site.
Table 2.2 Resources needed for evaluation

<table>
<thead>
<tr>
<th>Human resources</th>
<th>Financial resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal evaluation</td>
<td>Use of existing human resources</td>
</tr>
<tr>
<td>External evaluation</td>
<td>External consultants</td>
</tr>
<tr>
<td></td>
<td>International partners</td>
</tr>
<tr>
<td></td>
<td>ECDC</td>
</tr>
<tr>
<td></td>
<td>WHO</td>
</tr>
<tr>
<td></td>
<td>Travel costs</td>
</tr>
<tr>
<td></td>
<td>Consultant fees</td>
</tr>
</tbody>
</table>

**Olympic Lessons**

The human resource requirements of syndromic surveillance during a mass gathering are greater than during other times. The analysis, interpretation and reporting of data is often at a higher frequency and resources are often required over weekends and public holidays. It is therefore advisable, when assessing requirements, to plan for the additional short-term human resources that would be required for the mass gathering including:

- Additional data analysts
- Complete medical/senior epidemiologist cover
- Sufficient cover within the reporting team to allow out-of-hours working

**2.6 References**


3.1.1 Introduction

The specific data sources used for syndromic surveillance will vary depending on the setting (country or region), the health care structure and the resources available. In general, syndromic surveillance systems are opportunistic (i.e., they take advantage of data sources that are readily available).

In deciding which data sources may be suitable for use in a syndromic surveillance system, the first step is to consider its purpose. For example, the data source for a system whose main purpose is early warning will need to be one that collects data early in the onset of the disease under surveillance (or earlier than other surveillance systems); a system designed to monitor the occurrence of severe disease will need to use as a data source a health care setting where patients present with severe illness (for example, emergency departments or intensive care units).

Figure 3.1 Health pyramid showing possible data sources at different stages of health care-seeking behaviour

Adapted from the Kaiser Permanente Health Pyramid
In the first instance, to decide what the best data sources are, it is recommended that you consider the following basic questions and document the answers to help shape the design of your system (and inform evaluation at a later stage):

- What is the purpose of the system?
- Is it part of a national/regional surveillance programme for a specific event?
- What do you want to monitor with the system? Is it multiple or single conditions, infectious disease and/or environmental event?
- Will the system need to provide early warning or an indication of early impact assessment?
- Does the system need to report on a frequent basis?

For more information on data source evaluations see Triple-S Guidelines for the Assessment of Data Sources (Ziemann et al., 2013).

Olympic Lessons

It is important to identify in advance of a mass gathering any new data sources that will be required. An option appraisal should be undertaken to determine:

- Key areas of the national health care service that visitors will be encouraged to use during the mass gathering, for example, emergency departments that are NOT currently part of your existing syndromic surveillance system
- Any changes, or additions, to the health care service that may have implications for your existing syndromic surveillance system

It is also important to consider legacy in the development of new systems. Try to facilitate the continuation of new systems after the mass gathering.

3.1.2 Range of data sources

This section presents a range of possible data sources. It includes their strengths and weaknesses, the different variables that can be collected and the methods for collecting the data.

In determining which of the available data sources may be appropriate for your system, it is useful to have an understanding of the many and varied types of syndromic surveillance systems that are currently used in European countries.

Figure 3.2 illustrates the progression of useful data sources by underlying cause and associated health-seeking behaviour. Some data sources are used by both human and animal syndromic surveillance systems, for example:

- health professionals (for example, GPs for human health, veterinarians for animal health)
- telephone health lines
- drug sales
- internet activity.
The types of human and animal health data source are summarised below, and Table 3.1 provides information on each source’s strengths and weaknesses. For more information on existing data sources in Europe see the Triple-S projects human and veterinary inventory reports available at www.syndromicsurveillance.eu.

- General practitioners (GP) or primary care doctors: data will reflect the clinical diagnoses made by the GPs for a range of medical conditions and infections (see Example 3.1).
- Emergency departments: data on numbers of visits and diagnoses (or persons presenting with given symptoms) are collected in addition to the final discharge status of the patient. These systems are useful in monitoring more severe diseases/conditions, as persons with such conditions are more likely to present to emergency departments (see Example 3.2).
- Emergency medical services/ambulance dispatch activity: such systems include information on the presenting symptoms and severity of patients (see Example 3.3).
- Telephone health (telehealth) lines: the underlying reason for the call (for example, presenting symptom or indication of the clinical algorithm used to triage the caller) can be used as the syndromic indicator (see Example 3.4).

Figure 3.2 Human and animal health data sources by underlying infection and associated behaviours

Adapted from Dupuy, 2012.

NOTE: although some syndromic surveillance systems in Europe also use human mortality data (all cause deaths or underlying cause of death), the use of such data is the focus of EuroMomo project (European Mortality Monitoring Project www.euromomo.eu) and will not be covered in detail in these guidelines.
EXAMPLE 3.1 The Health Protection Agency/QSurveillance GP surveillance system
Country: United Kingdom
Data source: GP consultations
Status: active (2004)

This is a large national GP surveillance network covering approximately 30% of the population. The system was primarily developed for pandemic influenza, however it is able to monitor 30 clinical indicators (for example influenza-like illness, diarrhoea and vomiting). GPs are not required to undertake any specific data recording as collection is automated.

Data are routinely captured, analysed and published on a weekly basis. However, during a serious incident, the frequency of data capture can be increased to daily. The system assists the HPA and the Department of Health by providing community-based morbidity data during routine seasonal outbreaks of disease (for example, influenza, norovirus) and also during national incidents (for example, volcanic ash clouds).

Routine outputs include incidence rates per 100,000 population for a number of indicators. Tabulated data for key indicators are also provided to local health authorities.

Example of weekly output from the HPA/QSurveillance system for influenza-like illness


1. On 1 April 2013 the Health Protection Agency transferred its functions into Public Health England.
2. On 1 April 2013 the HPA/QSurveillance GP system was changed to the Public Health England GP surveillance system.
EXAMPLE 3.2 Syndromic surveillance system SurSaUD®/Emergency department network OSCOUR®
Country: France
Data source: Emergency department attendances
Status: active (2004)

SurSaUD® combines four data sources, one of which is the OSCOUR® network which collects data from emergency departments. The network was set up in 2004, following the heat wave the previous year, by the French Institute for Public Health Surveillance (InVS).

For an emergency department to participate, its data collection system must be computerised. In 2012, more than 370 of the 650 emergency departments were involved in the system, covering 60% of visits (approximately 25,000 adults and 3,500 child visits are analysed each day). All French regions are covered, including the overseas territories.

The network helps InVS to: (1) improve health surveillance of the population; (2) provide early warning of expected and unexpected or unusual health events in infectious diseases and environmental health; (3) follow-up and assess the public health impact of such events; and (4) provide objective intelligence to inform decision/policy makers.

Assessment of the heat indicator (dehydration, hyperthermia, malaise, hyponatremia) during summer 2006

Source: Data taken from Josseran et al., 2009.

Measles outbreak (August 2009 – August 2011)

Figure taken from the OSCOUR® Weekly Bulletin.
The European syndromic surveillance project SIDARTHₜₐ is currently active in the emergency medical dispatch centre of the state of Tyrol, Austria and in the university hospital emergency department in Santander, Cantabria, Spain.

It automatically analyses routinely collects emergency medical dispatch data on a daily basis, displays the results on a dashboard (see Figure) and sends a report by email to the regional public health authority.

The system analyses syndromes based on Advanced Medical Priority Dispatch (AMPDS) codes (for example, respiratory, gastrointestinal or cardiovascular problems). The algorithms that are currently used to detect statistical aberrations are EARS and CUSUM (for Poisson and normal distributed data).

The system is used as an additional and timely information source for early detection and reassurance.

**Daily emergency medical dispatch data displayed on the SIDARTHₜₐ dashboard, Austria**

*Source: www.sidartha.eu*
EXAMPLE 3.4 The HPA/NHS Direct Syndromic Surveillance System

Country: England and Wales
Data source: Telehealth line
Status: Active (1999)

NHS Direct is a multi-channel health advice and information service for the population of England. In the telephone channel, operators use a series of clinical assessment algorithms to evaluate the symptoms of each patient and the predominant symptom is recorded.

The service is nurse led, other health professionals — such as pharmacists — are used where appropriate. Callers are triaged and their reported symptoms and the severity are evaluated to produce recommended call outcomes that include: advice for self-care, referral to an emergency department, referral to urgent general practitioner care, or referral to routine general practitioner care. All outcomes are dependent on the seriousness of the call and the ‘risk status’ of the caller — i.e. whether they are young or old, or have other illnesses.

The service is accessible all day, every day and provides a reliable and continuous feed of data that are used by the Health Protection Agency to form the basis of the HPA/NHS Direct syndromic surveillance system. A number of key syndromes are monitored, for example, cold/flu, vomiting, diarrhoea, which are indicative of commonly circulating pathogens for example influenza, norovirus, rotavirus.

The daily percentage of total calls is monitored, this helps to smooth the data as the number of calls fluctuates over weekends and public holidays. NHS Direct traditionally provides early warning of approaching influenza and norovirus epidemics: a patient might contact NHS Direct before seeking a GP consultation or visiting an emergency department. The NHS Direct system has also been used in the surveillance of the impact of heat waves.

Diarrhoea and vomiting calls to NHS Direct compared to baseline

‘Speedometer’ dials

These dials show call levels versus expected levels or thresholds for a range of syndromic indicators. They provide a visually attractive and user-friendly presentation of data.


1. From 1 April 2013 the Health Protection Agency transferred its functions into Public Health England.
The internet is only used as a data source by a few syndromic surveillance systems in Europe. However, its use is becoming increasingly common as a result of changes in the health-seeking behaviour of the population and the development of health websites.

- Health websites: similar to telehealth, these systems monitor the health-seeking behaviour of the population through websites that provide health advice (see Example 3.5). There are also examples of national web portal services that are used by both the general public and by medical professionals, for example, Health Gate in Finland. Data can be analysed in a similar fashion to the telehealth system described above.

- Queries made on internet search engines: can be monitored to track population-based internet health enquiries. These data can be regarded as proxies for the surveillance of, for example, community-based infectious disease activity. While syndromic surveillance systems monitoring general internet search engine queries are not yet common, Google Flu Trends, for example, have established their potential (see Example 3.6).

- Prescription drug sales: can be used as a proxy for the surveillance of, for example, infectious disease (see Example 3.7).

- Over-the-counter-drug sales: can be used to monitor the self-treating behaviour of the population prior to the consultation with a health care professional. Such systems generally monitor numbers of sales of individual drugs or groups of drugs that are relevant to particular conditions such as anti-diarrhoeal drugs and rehydration therapy for monitoring gastrointestinal infections.

- Absenteeism data: a measure of absenteeism rates from a variety of sources including workplace and school. The number of cases of absenteeism and the underlying reason for the absence (if available) is used.

- Social media activity: syndromic surveillance systems that use social networking sites such as Facebook and internet forums have not yet become common, and those that do exist are still research projects. Such systems monitor activity across social network sites by looking for common text strings/phrases that are suggestive of increasing levels of self-reported illness in the community.

- Self-reported symptoms: websites where members of the public can register and keep a diary of symptoms, typically geared towards influenza. Examples include www.fluwatch.co.uk (UK), www.degrotegriepmeting.nl (NL) and www.influensakoll.se (SE).

Other examples of data sources in Europe include facilities such as ASTER (a military epidemiological surveillance system) and health services in immigration centres (system for epidemic-prone diseases set up following increased migration flows, Italy).
EXAMPLE 3.5 Get Well

Country: Sweden
Data source: Queries on Swedish medical website
Status: Active (2009)

Get Well is a syndromic surveillance system based on anonymous query data collected from a Swedish medical website. The system was developed at the Swedish Institute for Communicable Disease Control and is routinely used for syndromic surveillance.

It consists of two conceptual parts: one allows for trends to be extracted, based on user specified requests; the other permits tailored analyses of particular diseases, where more complex statistical methods are applied to the data. Examples of the latter are automated analyses of influenza and norovirus reported on a weekly basis.

A system based on web queries is flexible: it can be adapted to any disease; it can receive information on individuals other than those who seek medical care; and the data do not suffer from reporting delays as they are in near real-time.

An example of an influenza graph from Get Well (2012)

The web query-based influenza surveillance is published weekly on the internet (in Swedish):
Further information is available at: http://smi.se/publikationer/veckorapporter/webbsok
Google provides two tools that can be used to build a syndromic surveillance system: Google Trends and Google Correlate.

Google Trends (http://www.google.com/trends/) allows the popularity of specific search terms to be extracted. This can be, in principle, restricted to a geographical area. Restriction can also be based on the language of the search term if it is specific enough (such as Nordic and some East European languages).

In Google Correlate (http://www.google.com/trends/correlate/), useful search terms (identified as proxy measures for surveillance) can be identified by providing time series such as one from laboratory-based surveillance in the case of flu surveillance.

Google Trends screenshot showing the temporal and spatial distribution of the flu epidemic in France in 2012–13 compared with the previous years.
Veterinary clinics: data based on the clinical diagnoses made by veterinarians for a range of medical conditions (see Example 3.8).

Slaughterhouses: meat inspection is conducted in every European slaughterhouse to ensure the food safety of meat products. Inspection can produce valuable indicators for syndromic surveillance (for example, condemnation rates, the frequency of reasons for condemnation). Slaughterhouses are particularly interesting since they represent the final destination of most livestock and so have high population coverage. At present there are two pilot systems using slaughterhouses data for syndromic surveillance in Europe: Nergal-Abattoir in France (see Example 3.9) and NASEVA in Finland.

EXAMPLE 3.7 Scottish Syndromic Surveillance System
Country: UK (Scotland)
Data source: Electronic prescription data
Status: Active (2009)

The NHS ePharmacy Programme introduced the ‘Electronic Transmission of Prescriptions’ and supports improvements in the end-to-end prescribing, dispensing and payment processing of prescriptions.

The electronical prescription data is also used in syndromic surveillance: during the Influenza pandemic antiviral prescription for all Scotland was analysed to monitor the trend of the pandemic.

Antiviral prescriptions (Oseltamivir and Zanamivir) Scotland (April 2008 – January 2009)

Source: Health Protection Scotland (2010).
EXAMPLE 3.8 VetCompass (Veterinary Companion Animal Surveillance System)
Country: United Kingdom
Data source: Veterinary clinics
Status: Active (2007)

The Royal Veterinary College, in collaboration with the University of Sydney, is undertaking a long-term nationwide survey of diseases among companion animals (dogs, cats, etc.).

The aims of this project are to investigate the range and frequency of the health problems seen in companion animals by veterinary surgeons working in general practice in the United Kingdom, and to highlight the major risk factors for these conditions.

They are doing this through the routine capture of first opinion on clinical data (first diagnosis hypothesis) via electronic animal records held with Practice Management Systems – the software commonly used in veterinary clinics.

Further information is available at: www.rvc.ac.uk/VetCompass/Index.cfm).

EXAMPLE 3.9 NERGAL-Abattoir
Country: France
Data source: Slaughterhouses
Status: Pilot phase

NERGAL-Abattoir is a pilot database created in 2005 to collect data from ten bovine abattoirs in real time during the normal slaughtering process. A study is being conducted from 2011 to 2014 to evaluate the usefulness of these data in implementing a syndromic surveillance system. Ideas for improving the pilot system will be taken into account for the future national system, which will collect real-time data from all abattoirs in France.

Example of outputs for farmers from the NERGAL-Abattoir

![Graphs showing data analysis for NERGAL-Abattoir](image)
Laboratories: when the first symptoms occur in animals, veterinarians make a list of suspected diagnoses and request laboratory tests to confirm the hypothesis. By monitoring the number and type of requests for testing, emergence of diseases can be detected earlier. Some pilot systems currently use these data for syndromic surveillance (for example, the Animal Health System in Switzerland and Farmfile in UK).

Farms or professional organisations: breeders or breeders’ associations usually collect production data for monitoring their activities (for example, milk production level and quality, fertility). Since a drop in production can be an early sign of disease, production indicators may be suitable for use in syndromic surveillance systems. In Europe some systems use such data, in most cases in addition to other data sources (see Example 3.10).

EXAMPLE 3.10 GMON (Health Monitoring System for Cattle)
Country: Austria
Data source: Veterinarian clinics, dairy cattle performance recording organisations
Status: Active (2006)

GMON is a widely used health monitoring system. Since 2006 veterinary diagnostic data (enforced mandatory collection) have been generated, validated and recorded in a centralised database. In addition to the provision of reports for herd management and preventive measures, the project’s objectives include the assessment of breeding values for health traits and the monitoring of health status.

Example of a report sent to farmers on reproduction, mastitis and metabolic problems in the herd

Drug sales/medications: the use of drugs in animal health is strictly regulated. Data collected for the surveillance of antibiotic prescriptions can be used for syndromic surveillance (Example 3.11).

Sales of over-the-counter drugs and preparations can be used for monitoring treatment of infectious diseases by the owner prior to consultation with a health care professional. In most cases, sales of individual drugs or groups of drugs for a specific condition are monitored (for example, anti-diarrhoeal drugs and rehydration therapy for monitoring gastrointestinal infections). There may be delays in obtaining these data.

Drugs sales after consultation with a health care professional could also be an important source of data if available in a timely manner.

EXAMPLE 3.11 VETSTAT
Country: Denmark
Data source: Drug sales
Status: Active (1997)

All data on the purchase of medicines (antibiotics and vaccines) for livestock are collected in Denmark. Data can easily be merged to look at the usage for, for example, specific animal species/disease syndromes/specific antibiotics within geographical areas or the entire country. The objective is to control the use of antibiotics. Farmers can compare their antimicrobial use to that at regional and local levels.

An example of outputs provided to farmers

• **Telephone health (telehealth) lines:** as in human surveillance, monitoring the number of callers to a telehealth service could be an important data source. In particular, the underlying reason for the call (for example, the presenting symptom or indication of the clinical algorithm used to triage the caller) can be used as the syndromic indicator (Example 3.12).

---

**EXAMPLE 3.12 GD Animal Health**
Country: The Netherlands  
Data source: Telephone helpline, rendering plants, production data  
Status: Active (2002)

In 2002, a telephone helpline for livestock was set up in the Netherlands. Farmers and veterinarians can contact the helpline and the data (on animal disease, symptoms or syndromes) from their calls are collected in a database. Census data from other sources are also collected: rendering plant, Identification and registration system, breeding organizations, milk quality data, milk production data, farm voluntary health certification statuses, laboratory test results.

Based on these different data, statistical analyses are carried out, and the outputs are discussed by an expert group to interpret, alert and determine what investigations are necessary. Quarterly reports for Government, levy boards and industry are produced.

For further information see: www.gddeventer.com and Bartels CJM et al., 2006.

---

• **Volunteer self-reporting:** are based on a network of participants who report cases on a voluntary basis (see Example 3.13). These participants can be a restricted population (for example, hunters reporting to the SAGIR system – surveillance of wildlife in France – when they find dead or diseased animals) or, in the case of influenza in wild birds, the general public (for example, reporting of abnormal mortality in wild birds to veterinary services).

---

**EXAMPLE 3.13 REPAMO**
Country: France  
Data source: Network of correspondents  
Status: Active (1992)

REPAMO, created in 1992, is the French surveillance system dedicated to disease in wild and farmed marine molluscs. It is run by Ifremer on behalf of the Ministry of Agriculture and has 19 correspondents in 13 locations on the Channel, Atlantic and Mediterranean coasts of France. Data are collected from local authorities, laboratories and REPAMO correspondents.

Its original objectives were to perform surveillance of notifiable diseases, investigate mortality outbreaks, and implement a 2–3-year survey on specific host-pathogen associations. Its ultimate goals are to prevent the introduction and spread of infectious agents, in particular, pathogens involved in notifiable diseases, and to monitor the changes in the pathogens already present on the French coasts.

For further information see: Guichard B et al., 2011.
• Rendering plants: according to European legislation, ‘fallen stock’ (dead animals) represents a risk to public health and must be rendered in authorised plants (with certain exceptions such as remote areas). The legislation also specifies that any person (public or private companies) collecting and rendering animal cadavers has to record certain information (quantity collected, identification number, date and place of collection). Many rendering plants collect additional information (weight of cadavers, production type) for their own purposes or because of specific national regulations.

Several European projects, aiming at monitoring disposal requests made by farmers, are currently being piloted (see Example 3.14).

EXAMPLE 3.14 OMAR
Country: France
Data source: Rendering plants
Status: Pilot phase

Example output from the OMAR pilot system

The OMAR project (Observatoire de la Mortalité des Animaux de Rente) was launched in 2009. Its aim is to analyse the data collected by companies that work with fallen stock and design a monitoring system for detecting anomalies that could be associated with health events.

Currently, only retrospective analyses are conducted and the usefulness of the data for syndromic surveillance is still being evaluated.

Calls made by farmers to rendering plants are registered on a daily basis (including the number of animals, species, age group, farm location, date of call) and automatically transmitted to the system. About 1.2 million cattle death notifications are collected each year.

For further information see: Perrin J-B et al. (2010).
• **National registers:** represent another important source of data on livestock mortality and reproduction (calving interval). For example, all member states have been required to set up a computerised database for the identification and registration of cattle. Cattle owners have to report within seven days to their National Cattle Register, all movements to and from the holding, and all births and deaths of animals on the holding, along with the dates of these events. These registers are supposed to be comprehensive and rapidly updated, and so may be suitable for implementing mortality surveillance systems.

A retrospective study carried out using the national registers has demonstrated the usefulness of these data for estimating the impact of epidemics (Perrin et al., 2010).

Table 3.1 presents an exhaustive list of the existing sources identified in the Triple-S inventory, including the strengths and weaknesses of each.
Table 3.1 Syndromic surveillance data sources and their main characteristics

<table>
<thead>
<tr>
<th>Data source Description</th>
<th>To do what?</th>
<th>+ Strengths</th>
<th>– Weaknesses</th>
<th>Human examples in Europe</th>
<th>Veterinary examples in Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>General practitioner/veterinarian consultations Provisional clinical diagnoses</td>
<td>Monitor community levels of medical conditions</td>
<td>+ Routinely generated</td>
<td>– Different GPs’ coding systems</td>
<td>Network for surveillance and control of communicable diseases (Cyprus)</td>
<td>Veterinary companion animal surveillance system – VetCompass</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Requirement in most countries</td>
<td>– Diversity of software used</td>
<td>Medical on-call service for ILI surveillance – DMOS (Denmark)</td>
<td>Register for cattle farms – NASEVA (Finland)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Diagnoses made by health care professional</td>
<td>– Diagnosis at clinical discretion which is subjective</td>
<td>Real time alert and surveillance – ASTER (France)</td>
<td>Register for swine farms – SIKA (Finland)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Widely available in electronic format</td>
<td></td>
<td>GP network – Réseau SOS Médecins – SurSaUD® (France)</td>
<td>Poultry practice data (UK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HPA/Qsurveillance National Surveillance System (UK England and Wales)</td>
<td>Small Animal Veterinary Surveillance Network – SAVSNET (UK)</td>
</tr>
</tbody>
</table>
Table 3.1 Syndromic surveillance data sources and their main characteristics (continued)

<table>
<thead>
<tr>
<th>Data source Description</th>
<th>To do what?</th>
<th>+ Strengths</th>
<th>– Weaknesses</th>
<th>Human examples in Europe</th>
<th>Veterinary examples in Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency departments</td>
<td>Monitor community levels of medical conditions</td>
<td>+ Widely available in electronic format</td>
<td>– Different coding systems</td>
<td>ED network OSCOUR® from SurSaUD® (France)</td>
<td></td>
</tr>
<tr>
<td>Emergency department attendances with diagnosis data</td>
<td>Provide early warning of outbreaks of severe disease, for example Haemolytic Uremic syndrome</td>
<td>– Lack of specific coding within Emergency department</td>
<td></td>
<td>Emergency Department Syndromic Surveillance System – EDSSS (England, UK)</td>
<td></td>
</tr>
<tr>
<td>Emergency dispatch (ambulances)</td>
<td>Monitor patient-reported signs and symptoms</td>
<td>+ Real-time</td>
<td>+ Early warning of localised incidents</td>
<td>European Emergency Data-based syndromic surveillance System (SiDARTHa, Europe)</td>
<td></td>
</tr>
<tr>
<td>Emergency medical dispatch data</td>
<td>– Lack of uniform systems at a national and international level</td>
<td></td>
<td></td>
<td>BIOALARM (Denmark)</td>
<td></td>
</tr>
<tr>
<td>Laboratories Test requests and sales</td>
<td>Monitor laboratory sales and test requests as a proxy of early onset</td>
<td>+ Commonly registered</td>
<td>+ Electronic format</td>
<td>G8 Summit surveillance system (Scotland, UK)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Lack of specificity</td>
<td></td>
<td>Animal Health system (Switzerland)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Defra equine surveillance reports (UK)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Farmfile (UK)</td>
<td></td>
</tr>
<tr>
<td>Data source Description</td>
<td>To do what?</td>
<td>+ Strengths</td>
<td>– Weaknesses</td>
<td>Human examples in Europe</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td>Health websites</td>
<td>Provide a non-medical diagnosis of patient-reported signs and symptoms with recommendations for further care/treatment</td>
<td>+ Ease of access</td>
<td>– Participation rates, excludes non-web users</td>
<td>FluSurvey (UK)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Data in electronic format</td>
<td>– Lack of specificity of data;</td>
<td>Generating Epidemiological Trends from Web Logs – GETWELL (Sweden)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Flexibility in designing service/questions</td>
<td></td>
<td>Like (Sweden)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Ability to monitor true community-based illness</td>
<td></td>
<td>Health Gate (Finland)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GrippeWeb (Germany)</td>
<td></td>
</tr>
<tr>
<td>Absenteeism</td>
<td>Monitor levels of sickness in schools/nurseries and/or workplace</td>
<td>+ Clearly defined/known population at risk that can be used as denominator</td>
<td>– Not routinely available in real time</td>
<td>No known established systems, but examples of pilot work in the UK during the 2009 influenza A(H1N1) pandemic as a tool for early detection of influenza activity in the community</td>
<td></td>
</tr>
<tr>
<td>Absenteeism rates in schools/nurseries or workplace</td>
<td></td>
<td>+ Enables the creation of incidence and prevalence rates</td>
<td>– Lack of underlying cause of sickness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>farms and professional organizations</td>
<td>Monitor herds' health</td>
<td>+ Commonly registered for livestock</td>
<td>– Lack of specificity</td>
<td>Health monitoring system for cattle – GMON (Austria)</td>
<td></td>
</tr>
<tr>
<td>Production data</td>
<td>Detect the emergence of new diseases</td>
<td></td>
<td></td>
<td>NASEVA (Finland)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Animal Health system (Switzerland)</td>
<td></td>
</tr>
<tr>
<td>Data source Description</td>
<td>To do what?</td>
<td>+ Strengths</td>
<td>– Weaknesses</td>
<td>Human examples in Europe</td>
<td>Veterinary examples in Europe</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Immigration centres</td>
<td>Monitor the health status of immigrant populations and respond rapidly to any health emergency</td>
<td>+ Primary source of timely health data for this population at a national level</td>
<td>– Uncertainty about the total number of immigrants in migration centres at any given time</td>
<td>Syndromic surveillance of epidemic-prone diseases in response to an exceptional influx of immigrants from north Africa to Italy (Italy)</td>
<td></td>
</tr>
<tr>
<td>Health desks in immigration centres</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telephone helplines</td>
<td>Record patient-reported signs and symptoms</td>
<td>+ Routinely generated and available in real time</td>
<td>– Only for mild illness/disease</td>
<td>NHS Direct (England, UK)</td>
<td></td>
</tr>
<tr>
<td>Use of telephone health lines to monitor reported illness</td>
<td>Record information from veterinarians and farmers about signs and symptoms in animals</td>
<td>+ Electronic format</td>
<td></td>
<td>NHS24 (Scotland, UK)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Earlier warning</td>
<td></td>
<td>GD Animal health (The Netherlands)</td>
<td></td>
</tr>
<tr>
<td>Slaughterhouses</td>
<td>Monitor different types of condemnation causes, proportions of condemnations, etc.</td>
<td>+ Information not otherwise available</td>
<td>– Final destination of most livestock</td>
<td>NASEVA (Finland)</td>
<td></td>
</tr>
<tr>
<td>Data on carcass condemnations recorded by veterinary services</td>
<td></td>
<td>+ Final destination of most livestock</td>
<td>– Difficult to register in real-time.</td>
<td>SIKAVA (Finland)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Not often in electronic format yet, but more pilot databases are being implemented</td>
<td></td>
<td>Nergal-Abattoir (France)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Animal Health system (Switzerland)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Innova AM and PM (UK)</td>
<td></td>
</tr>
<tr>
<td>Data source Description</td>
<td>To do what?</td>
<td>+ Strengths</td>
<td>– Weaknesses</td>
<td>Human examples in Europe</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| Rendering plants       | Monitor livestock mortality | + Exhaustive information  
+ Rapidly available  
+ Data collection is mandatory according to European legislation | – Lack of specificity  
– Data are not individual for all species | OMAR (France)  
PROVIMER (Spain)  
Dairy cattle mortality – MBL (Italy)  
Over 48 month fallen stock (UK) |
| Social networking       | Provide true community estimates of disease/illness as it includes individuals who would not normally seek medical care or farmers who would not ask for information from veterinarians for animal health | + Benefits from current popularity of social networking  
+ Reflects changing structure of health-seeking behaviour | – Lack of specificity of data  
– Costs of accessing/purchasing data  
– Excludes those non-users of social networks |
Once you have decided on the purpose(s), data sources and data providers, you must decide which variables to collect.

The main variables collected for syndromic surveillance are those relating to health, yet in most cases other variables are collected and are an important part of the system’s output.

The availability of variables that fulfil the requirements of the purpose you have defined for your system should be assessed with data providers at an early stage. Generally, the data providers themselves will provide a list of the variables that they normally collect or that it is possible to collect.

Each variable has a different purpose in providing information of use to the system. These variables may include patient demographics and the details of the signs/symptoms/provisional diagnoses made during an event, referral path and severity of illness.

At this stage, it is advisable to also get advice from data analysts/statisticians about which variables are the most suitable, and on the format required to allow appropriate analyses and meaningful interpretation of the data.

Once you have decided on the protocol of data analysis and its required variables, you should initiate the collection of the data with the data providers. The list below includes a range of variables that can be collected from a data source.

### 3.2.1 Signs, symptoms and proxy measures

The main variables of a syndromic surveillance system are the health variables chosen for surveillance, which can be:

- Unspecific clinical data. For example, sign/symptom, syndrome (a group of signs/symptoms), chief complaint, provisional diagnosis (made by GP, veterinarian, farmer).

- A proxy measure of population health. For example, drug sales, absenteeism, web queries, carcass condemnation rates, number of laboratory submission requests.

This information provides one of the most critical characteristics of a syndromic surveillance system. Therefore an assessment of the availability of these variables within the data source is vital.

In evaluating the range of variables within a data source, you must keep in mind the main purposes of the system and tailor the collection of variables (symptoms, diagnoses, proxy data) to meet these needs. For instance, do I need specific coding, or will generic reporting be sufficient (for example, respiratory disease or acute bronchitis)?
3.2.2 Demographics

The data source should cover as much information as possible on the target population to ensure a level of representativeness. Consideration must be made as to whether the data source is, as far as possible, comparable to national/regional statistics with regard to such variables as age, sex, ethnicity and deprivation.

To assess the demographic coverage of the data source, a descriptive analysis is necessary. It is also necessary to consider any changes in the coverage of the population (for example, during the summer in tourist areas or variations in movements between farms and slaughterhouses due to external factors such as the price of meat). These issues can be critical when monitoring disease outbreaks or responding to other health incidents (heat wave, mass gathering).

Age Collecting the variable ‘age’ (or age groups) can be useful in identifying outbreaks and alerts, and in monitoring the progression of an epidemic that affect specific age groups. It can also be useful in detecting changes in the normal age distribution of a disease or incident. For instance, this was of particular note in the 2009 H1N1 pandemic where the expected age groups affected were much younger and not necessarily at-risk of developing flu.

However, when using this variable, confidentiality issues should not be neglected, as specific age taken together with other variables could allow an individual to be identified. For this reason, instead of recording ‘date of birth’, it may be preferable to collect information on ‘age in years’ or ‘year of birth’. For animal health, information on age of animals is not treated as sensitive information.

Sex The variable ‘sex’ can be a useful source of information as it can provide additional information on the makeup of an outbreak. In the event of unusual outbreaks of disease (for example E. coli 0157/ HUS (Haemolytic Uremic Syndrome), information on sex can provide useful evidence to orient an outbreak investigation.

In animal health, information on sex may need to include information on the neutered status or the parity (cows/heifers) of animals because this information could have an impact on the pathology.

Species, breed and production type (animal health) The population being targeted needs to be well defined in terms of information on the species, breed and production type.

Knowing the species of animal affected by a disease or event is important because diseases can be specific to one or more species (for example, rabies can affect all mammals, West Nile can affect equine and birds, atrophic rhinitis of swine affects only this species).

Moreover, within individual species, the sensitivity to a disease can vary depending on the breed of animal, especially in dogs and cats.
For livestock, production is related to breeding practices and thus to certain pathologies, for example, mastitis or lameness are most frequent among dairy cattle.

Variables recording ‘animal breed’ and ‘production type’ are available in most animal health data sources, but often the amount of missing information for these variables can hinder their use for syndromic surveillance purposes. A full assessment of the completeness of recorded information for these variables should be carried out before determining their inclusion in the surveillance definitions.

### 3.2.3 Denominator

It is important to identify the population under surveillance, but information on the ‘denominator’ population is only necessary if rates are to be calculated. Initially, this issue should be discussed with data analysts/statisticians and/or senior epidemiologists to decide on the reporting statistic from the system.

The denominator will differ for each data source, some examples of which are presented in Table 3.2.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Denominator</th>
<th>Example numerator</th>
<th>Example of reporting statistics</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP networks</td>
<td>Total registered population</td>
<td>Number of diarrhoea consultations</td>
<td>Rate per 100,000 population</td>
<td>Denominator easy to access through registered patient lists For animals, denominator easy to access for species with a national register. More difficult for other species</td>
</tr>
<tr>
<td>Emergency Department</td>
<td>Total number of admissions</td>
<td>Number of admissions for asthma attacks</td>
<td>Percentage of total admissions</td>
<td>Denominator easy to access through daily recorded admissions to emergency department</td>
</tr>
<tr>
<td>Telehealth</td>
<td>Total number of telephone calls</td>
<td>Number of calls about diarrhoea</td>
<td>Percentage of total calls</td>
<td>Total population can be a good proxy for this denominator as a large part of the population has a telephone</td>
</tr>
<tr>
<td>Websites</td>
<td>Total number of visits to a site</td>
<td>Number of visits to a page on a particular topic, e.g. influenza</td>
<td>Percentage of total page visits</td>
<td>Numerator difficult to determine due to potential duplicate visits. If IP is logged, duplicates can be filtered out to make the numerator slightly more reliable</td>
</tr>
</tbody>
</table>
3.2.4 Geographic information

It is vital to ascertain the geographical coverage of the data source to ensure that it covers the area that the syndromic surveillance system intends to cover and to determine, within the area of surveillance, the smallest area unit available. To do so the following questions should be considered:

- Is the data source/provider a national resource i.e. does the data source cover a country/state/local health area?
- Does the data source have sufficient coverage over these areas?
- Is the coverage sufficiently large to allow for epidemiological data to be analysed for a small geographical output area?
- Are the data from different regions comparable in terms of medical information and population structure?

If the syndromic surveillance system is a regional/state system, you should assess its compatibility with existing or planned systems in other regions/states so that the different systems can be used in tandem when monitoring nationwide incidents.

The most relevant epidemiological unit for farm animals’ health is the herd. However, the number of occurrences of a syndrome is low at this level and, depending on the population coverage of the data available, this may result in a small number of recorded cases per time point (for example, day). Monitoring low to sparse counts of syndrome occurrences may result in the need to group counts over a longer time scale, such as week or month, which will reduce the timeliness of the system. Alternatively, statistical methods for monitoring low or even rare events are available.

Another key issue, especially with regard to systems using GPs and emergency departments as data providers, is whether the system is meant to have exhaustive coverage or is to be a sentinel system. The advantage of exhaustive coverage (for example, including every emergency department in the area) is that it captures the whole population under surveillance. This enables measurement of the impact of a local event on morbidity (for instance, linking the emergency admissions with the source of a local chemical accident). A sentinel system, although representative of the area under surveillance, can only cover a smaller proportion of the population that may not have been particularly exposed to the risk under assessment.

To choose the proper scale for the syndromic surveillance system, geographical variables need to be collected. The following illustrates some examples:

- Post/zip code: The provision of a geographical code can help to identify the location of the case/patient/animal owner/herd, which can be important for mapping the origin and spread of incidents and/or outbreaks. The
post/zip code of the case is most commonly used. However, you must consider ethical and confidentiality issues of this practice, given that postcodes, when combined with other demographic data, can lead to the patient/herd's identification.

For this reason, the data provider often ‘truncates’ the post code, so that it cannot be used to identify an individual but is still useful as a geographical identifier for surveillance purposes. Confidentiality can also be respected by ensuring that only aggregated results are produced.

It is also important to have information on the stability of post/zip codes (i.e., how often they are changed or updated).

**Health district code**

In human health surveillance, an alternative to post codes are health district codes. These are particularly useful if the syndromic surveillance system collects aggregated data from the data source and thus is not able to identify individual case/patient post codes.

**IP address**

For internet-based surveillance systems, an IP code can be used to locate a user by identifying the computer being used to make the web query. However, large service providers or national organisations may appear to be in a single location, greatly limiting the usefulness of this method.

**Data provider information**

Sometimes the only geographical identifier will be the location of the data provider (GP/care centre/emergency department).

When analysing data, if the data on cases are aggregated by, for example, age, sex, or region, then the denominator must also be broken down in the same way. For instance, if reporting on GP consultation rates per 100,000 of influenza-like illness by age group, information on the age, sex, and region of the denominator population needs to be available.

### 3.2.5 Temporal information

A date stamp is required to confirm the date of attendance and to enable accurate reporting from a disease database (for example, date of the visit/call/query).

If available, a time stamp indicating the time of day that the event occurred (for example, symptom onset) can also be useful. In fact, if a major public health incident occurred, it would be possible to report cases within a day or even more rapidly (for example, on a four-hourly basis) and also to identify unusual times of presentation during the day, which could indicate the development of an incident. See Example 3.15).
EXAMPLE 3.15 The Emergency Department Syndromic Surveillance System (EDSSS)
Country: England
Data source: Emergency department attendances
Status: Active (2011)

In the Emergency Department Syndromic Surveillance System (EDSSS) each emergency department attendance has a ‘date/time stamp’ (date and time of attendance are recorded). Although data are only collected and analysed on a daily basis, it can be useful to compare attendance trends over a 24-hour period to identify when peak activity occurs. Looking at these data by day can also reveal the busy periods in the emergency department.

Although hourly data are not routinely used, if a serious public health incident did occur, they could be analysed to identify the time when increased numbers of individuals began to present.

Emergency department daily attendances (percentage by hour of day)

Source: Elliot et al., 2012.

3.2.6 Information on severity

Indication of the severity of a symptom can be obtained through patient discharge status, triage assessment and call disposition (in telehealth). See Example 3.16.

This information can enhance a syndromic surveillance system by providing an indication of the severity of a public health event. When routinely monitored, it is possible to assess deviations from the norm, which would suggest possible changes in the severity of the presenting health condition.
‘Severity indicators’ measure a change in the outcome of calls/consultations that suggest a more severe illness, for example, more of the visits for acute respiratory illness to emergency departments resulting in admission to high dependency care. These may be an important part of mass gathering surveillance as they can provide an alert to potential deliberate attacks/emerging diseases. It is therefore important to discuss the possibility of developing severity indicators across as key indicators for syndromic systems.
3.2.7 Treatment/prescriptions

Depending on the data source, it may be possible to obtain information on treatment or drugs prescribed. These events may be recorded individually, for example antiviral prescriptions, or may be linked to a disease event, for example antiviral prescriptions during an influenza-like illness (see Example 3.17).

These data can also be used to monitor changes in the severity of an evolving incident/outbreak with the principle that changes in treatments or the prescription of drugs may indicate more severely ill patients requiring medical interventions.

EXAMPLE 3.17 The Health Protection Agency/QSurveillance GP surveillance system
Country: United Kingdom
Data source: GP consultations
Status: active (2004)

The HPA/QSurveillance GP surveillance system is able to monitor aggregated counts of patients with a provisional diagnosis of influenza-like illness, as well as patients who have been prescribed an influenza antiviral. During the 2009 pandemic this prescribing indicator was used to monitor the number of patients presenting to GPs and being prescribed antivirals.

Influenza-like illness with antivirals prescribed, consultation rate per 100,000 population during 2009 (blue line) compared to 2008 (grey line)

Graph taken from the HPA/QSurveillance National Surveillance System Bulletin (www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/RealtimeSyndromicSurveillance).

1. On 1 April 2013 the Health Protection Agency transferred its functions to Public Health England and the HPA/QSurveillance GP system was changed to the Public Health England GP surveillance system.
3.2.8 Syndromic surveillance minimum dataset

The minimum set of variables considered to be essential in setting up a syndromic surveillance system are:

- Signs/symptoms and proxy
- Demographics (age), and for animal health, species (production type information is a bonus when available.)
- Geographical information
- Temporal information (date of call/visit)

Most syndromic surveillance systems in Europe collect the majority of the variables in the minimum dataset. Many systems collect additional variables to enhance syndromic surveillance (see Examples 3.18 and 3.19).

**EXAMPLE 3.18** The French Syndromic Surveillance system SurSaUD®

Country: France
Data source: Emergency department and emergency GP consultations

<table>
<thead>
<tr>
<th>Variables</th>
<th>ED network OSCOUR®</th>
<th>GP emergency associations/ SOS Médecins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms and proxy measures</td>
<td>Main and associated diagnoses (ICD-10)</td>
<td>From 1 to 3 chief complaints</td>
</tr>
<tr>
<td></td>
<td>Chief complaint (free text)</td>
<td>From 1 to 3 medical diagnoses</td>
</tr>
<tr>
<td>Demographics</td>
<td>Date of birth</td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>Sex</td>
</tr>
<tr>
<td>Geographic information</td>
<td>City of residence (name and post code)</td>
<td>City of the call (name and post code)</td>
</tr>
<tr>
<td></td>
<td>Emergency department and service (adult or paediatric)</td>
<td>City where SOS Association is located</td>
</tr>
<tr>
<td>Temporal information</td>
<td>Date and time of the visit</td>
<td>Date and time of the call</td>
</tr>
<tr>
<td></td>
<td>Date and time of the discharge</td>
<td></td>
</tr>
<tr>
<td>Severity information</td>
<td>Severity level (5 categories + death)</td>
<td></td>
</tr>
<tr>
<td>Other information</td>
<td>Transport (personal transport, ambulance)</td>
<td>Hospitalisation (Y/N)</td>
</tr>
<tr>
<td></td>
<td>Discharge status (home, hospitalisation, death)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>If hospitalised, information on the specific ward</td>
<td></td>
</tr>
</tbody>
</table>
EXAMPLE 3.19 OMAR, observatory of livestock mortality
Country: France
Data source: Rendering plants, Cattle national data bases
Status: Active (2006)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cattle national data bases</th>
<th>Rendering Plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms and proxy measures</td>
<td>Number of deaths</td>
<td>Number of pick-up calls</td>
</tr>
<tr>
<td>Demographics</td>
<td>Sex</td>
<td>Species</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>Sex</td>
</tr>
<tr>
<td></td>
<td>Breed</td>
<td>Age category</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Production type</td>
</tr>
<tr>
<td>Geographic information</td>
<td>Post code of herd location</td>
<td>Post code of herd location</td>
</tr>
<tr>
<td>Temporal information</td>
<td>Date of death</td>
<td>Date and time of the disposal request</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Date of the cadaver collection</td>
</tr>
<tr>
<td>Severity information</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Other information</td>
<td>Reason for (birth, death, trade)</td>
<td>Weight of cadavers</td>
</tr>
<tr>
<td></td>
<td>Date of cattle movements</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Still born</td>
<td></td>
</tr>
</tbody>
</table>

Once you have chosen the data source that best matches your system’s requirements and have agreed upon the variables to be collected, the next step is to group the records into syndromes. Valid, reliable and automatic classification of syndromes is an essential component of computerised epidemic detection systems.

In human health, less so in animal health, clinical data are regularly coded into clinical groups. These codes can facilitate or directly lead to syndrome definitions.

Coding systems are lists of the possible codes or values used to code the information about the presenting symptoms or the provisional diagnosis.

Because the data provider is likely to be already using a coding system, it will be necessary to evaluate its usefulness for syndromic surveillance. If it does not allow the analysis of syndromes/indicators,
there is little advantage to developing the syndromic surveillance system using this data source.

If the data provider uses a coding system that is unsuitable for your needs or does not use one, then a coding system will have to be adopted. In developing or adapting a coding system, it is recommended that you consider the format that is easiest for the data provider and requires the least additional work and/or resources. It is also recommended to use, where possible, a coding system that allows comparisons to be made with existing systems within the country or in Europe, so that outputs can be accurately compared and interpreted.

### 3.3.1 Existing coding systems

The data sources described in Section 3.1.2 use different coding systems. Each has its individual properties, advantages and limitations, described in Table 3.3.

<table>
<thead>
<tr>
<th>Coding system</th>
<th>Description</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>International coding systems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Snomed (WHO)</strong></td>
<td>Comprehensive clinical healthcare terminology</td>
<td>Internationally accepted coding: allows for comparability with other surveillance systems</td>
<td>Complex*</td>
</tr>
<tr>
<td><a href="http://www.ihtsdo.org/snomed-ct/">www.ihtsdo.org/snomed-ct/</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ICD-10 (WHO)</strong></td>
<td>Global health information standard for mortality and morbidity statistics</td>
<td>Internationally accepted coding: allows for comparability with other surveillance systems</td>
<td>Complex*</td>
</tr>
<tr>
<td><a href="http://www.who.int/classifications/icd/">www.who.int/classifications/icd/</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ICPC (WHO)</strong></td>
<td>‘Reason for encounter’ classification system for primary care or general practice</td>
<td>Allows for classification of the patient’s reason for encounter, the problems/diagnoses managed and interventions</td>
<td>Limited to primary care</td>
</tr>
<tr>
<td><a href="http://www.who.int/classifications/icd/adaptations/icpc2/">www.who.int/classifications/icd/adaptations/icpc2/</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: in complex coding systems data providers may use a subset of relevant codes and therefore the full extent of the coding list would not be available for syndromic surveillance.
Table 3.3 Main coding systems used in human health syndromic surveillance (continued)

<table>
<thead>
<tr>
<th>Coding system</th>
<th>Description</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country-specific coding systems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read</td>
<td>UK NHS medical thesaurus used by general practitioners</td>
<td>Hierarchical coding system with a wide range of detailed coding</td>
<td>Limited to UK primary care</td>
</tr>
<tr>
<td><a href="http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes">www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manchester triage system (MTS) Mackway-Jones K et al. 2006</td>
<td>Emergency department triage coding system</td>
<td>Standardised coding list – comparable across different emergency departments</td>
<td>Not used consistently (used extensively in UK but limited use in rest of Europe)</td>
</tr>
<tr>
<td>Unified Diagnostic Dataset (UDDA) <a href="http://secure.collemergencymed.ac.uk/Shop-Floor/Informatics/CEM%20Unified%20Diagnostic%20Dataset%20(UDDA)">http://secure.collemergencymed.ac.uk/Shop-Floor/Informatics/CEM%20Unified%20Diagnostic%20Dataset%20(UDDA)</a></td>
<td>Emergency medicine diagnostic dataset – developed by UK College of Emergency Medicine</td>
<td>Standardised coding mapping from a number of existing systems including SnoMed and ICD10</td>
<td>Currently in trial/pilot phase in a small number of UK emergency departments</td>
</tr>
<tr>
<td><strong>Bespoke data provider-specific coding systems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHS Direct</td>
<td>Telehealth algorithm-based symptom assessment tool</td>
<td>Normally tailored to the data source and therefore will accurately reflect the presenting symptoms</td>
<td>Only used in NHS Direct and so not compatible with other coding systems/providers</td>
</tr>
<tr>
<td>SOS Médecins</td>
<td>SOS Médecins’ general practitioners use four thesauruses to code diagnoses. Four other thesauruses are used for coding the primary complaints when patients call the associations</td>
<td>Standardised coding easily comparable to international coding such as ICD-10</td>
<td>Several coding systems limited to French GP emergency associations</td>
</tr>
</tbody>
</table>
In animal health there is a lack of harmonisation for clinical terms and few national coding systems. Some surveillance systems use their own coding system such as GMON in Austria, PROVIMER in Spain), SIKAVA in Finland (Table 3.4).

Table 3.4 Main coding systems used in animal health syndromic surveillance

<table>
<thead>
<tr>
<th>Coding system</th>
<th>Description</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country-specific coding systems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French lesion coding system</td>
<td><a href="http://agriculture.gouv.fr/IMG/pdf/dgaln20068139z.pdf">http://agriculture.gouv.fr/IMG/pdf/dgaln20068139z.pdf</a></td>
<td>Coding system of lesions for bovine slaughterhouses</td>
<td>Used in all French bovine slaughterhouses</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Based on a network of experts in meat inspection</td>
</tr>
<tr>
<td><strong>Bespoke data provider-specific coding systems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMON (Austria)</td>
<td>List of 65 diagnoses for cattle</td>
<td>No free text (scroll down list)</td>
<td>Valid only for this system</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Harmonisation within the system</td>
<td>Difficult for comparison between systems</td>
</tr>
</tbody>
</table>

3.3.2 Translation of codes (harmonisation of coding systems)

In some syndromic surveillance systems, data providers may use different coding systems within one data source (for example, different GPs may use different coding systems). This should not limit the integration of these data sources into a syndromic surveillance system. One solution is to translate the different coding systems into a single system, before or just after the integration of the data in the system (see Example 3.20).

**EXAMPLE 3.20** GPs’ emergency associations SOS Médecins

Country: France
Data source: Emergency GP consultations
Status: active (2006)

In the French SurSaUD® syndromic surveillance system, 59 GP emergency associations are involved. Within the different associations there are four ‘thesauruses’ to code the chief complaint and the diagnosis. To analyse the data the French Institute for Public Health Surveillance (InVS), responsible for the syndromic surveillance system, has created its own thesaurus for the chief complaint and another for diagnosis, to enable translation of different codes with the same meaning into a new InVS code.
3.3.3 Mapping of codes

One further note regarding coding formats is the need to map underlying codes to a syndromic indicator. For example, if the syndromic surveillance system needs to report ‘influenza-like illness’, the underlying coding system used by the data provider may permit several or many different codes, all of them relating to influenza-like illness, to be used by the front line clinicians working for the data provider. In this instance, a mapping exercise needs to be undertaken to ensure that all relevant codes are correctly allocated to a particular syndromic indicator (see Example 3.21).

These issues are more relevant where the syndromic surveillance system reports on more generic indicators (for example, all respiratory illness or all gastrointestinal illness), where there will be large numbers of underlying codes used by clinicians that can be mapped to these syndromes.

To develop accurate mapping tables it is crucial to have access to the totality of possible coding. This is particularly important when the data source uses a bespoke coding system.

The Unified Medical Language System (UMLS) allows the mapping or alignment of many coding systems and is a good generic tool for further exploring this issue (www.nlm.nih.gov/research/umls/).

EXAMPLE 3.21 The Emergency Department Syndromic Surveillance System (EDSSS)
Country: England
Data source: Emergency department attendances
Status: Active (2011)

Mapping of underlying SnoMed and ICD-10 codes to respiratory diseases.

Illustration of mapping of codes from different systems to a single syndromic indicator

<table>
<thead>
<tr>
<th>SnoMed</th>
<th>ICD10</th>
</tr>
</thead>
<tbody>
<tr>
<td>• ‘All respiratory conditions’ diagnosis section</td>
<td>• Chapter J00–99</td>
</tr>
<tr>
<td>• + searches for:</td>
<td>• R060</td>
</tr>
<tr>
<td>‘influenza’</td>
<td>• R042</td>
</tr>
<tr>
<td>‘pneumonia’</td>
<td>• R05</td>
</tr>
<tr>
<td>‘tonsil’</td>
<td>• R068</td>
</tr>
<tr>
<td>‘sinusitis’</td>
<td>• R091</td>
</tr>
<tr>
<td></td>
<td>• R092</td>
</tr>
<tr>
<td></td>
<td>• H653</td>
</tr>
<tr>
<td></td>
<td>• H668</td>
</tr>
</tbody>
</table>
3.4.1 Data recording

There are two categories of data extraction:

Without additional work
This is the ideal scenario. The data provider is able to incorporate the dataset collection into an existing reporting schedule without incurring any additional work (apart from the initial setup of the reporting schedule). This is the best way to ensure compliance of the data provider with the syndromic surveillance system.

With additional work
There may be scenarios where a bespoke data collection system is required to extract data from the data provider. In this instance, additional resources may be required to fund the development of a data collection system.

EXAMPLE 3.22 Syndromic surveillance system SurSaUD®/Emergency department network OSCOUR®
Country: France
Data source: Emergency department attendances
Status: Active (2004)

In 2004, it was possible to implement ED network OSCOUR® because the Ministry of Health decided to promote and organise the computerisation of emergency departments in France. The French Institute for Public Health Surveillance (InVS), the Ministry of Health and the French Scientific Society for Emergency Medicine produced a minimum dataset (called RPU) that included data for epidemiological surveillance. The computerisation was based on technical specifications that integrated this minimum dataset. At present, this dataset is part of all emergency department software packages produced by different IT providers. This allows epidemiological data to be collected automatically. So that data providers are not burdened with additional work, only computerized emergency departments can be part of the OSCOUR® network.

EXAMPLE 3.23 Genoa's syndromic surveillance system of respiratory infections, gastroenteritis, acute hepatitis, fever and other diseases
Country: Italy
Data source: Emergency department attendances
Status: Active (2007)

An example of a syndromic surveillance system that requires additional resources is that the city of Genoa (Italy). In this system, which collects data from two emergency departments, the data are coded using a bespoke coding system and not the International Classification of Diseases. For this reason, each admittance must be reviewed and the diagnoses recorded. This is extremely time consuming (over 400 records must be checked each day, which requires about three hours per day for the person responsible). This is feasible in the Genoa system because there are few emergency departments involved. However, in the case of an incident with a huge increase in activity or in a system involving a substantial number of data providers, such as a national system, consulting all the records would require such an amount of additional human resources and time that it would be practically impossible to implement.
3.4.2. Data transmission

Two of the most important features of the procedures for transmitting data are automation and low maintenance.

Ideally, the transmission of data from data providers to the host organisation should be fully automated. This is so that the providers are not required to perform any manual processing, which would increase their workload. This also improves the efficiency of the syndromic surveillance system, given that there are fewer opportunities for human error to occur in data transmission and less chance of missing transmissions. This is particularly relevant during an incident when data providers have less time to ensure quality of data collection and transmission, but epidemiologists from the syndromic surveillance system need the data to assess the health impacts of the incident.

The system for transmitting data should also require little maintenance. Although highly sophisticated transmission systems exist, if they are not reliable or require specialised support to maintain them, it will waste the data provider’s resources and may result in the syndromic surveillance system receiving negative feedback. Simple data transfer systems can be as effective as complex ones.

Olympic Lessons

It is important to discuss any changes to the data transmission process (for example, frequency) with data providers in advance of the mass gathering. This will allow time for the data providers to implement any changes and the surveillance team to change internal data capture processes.

Data format

All formats should be electronic. Data format simply refers to the ‘medium’ in which data are transferred to the host organisation (for example, in structured spreadsheets or in lists such as XML).

The main difference between the various formats is the extent to which they are structured:

- Free format data such as text files cannot ensure that every line is complete;
- Structured but not typed format such as CSV and spread sheets (Excel, Open Office Calc) cannot ensure that every column has just single type of data;
- Typed formats such as some XML and JSON derivatives allow for automatic checks of the data types and structure. This allows the data provider to perform automated validity checks;
- Self-sufficient formats such as SDMX include all relevant data as well as metadata. These are rarely useful for syndromic surveillance.

It is necessary to ensure that the data format adopted is commonly accepted and widely used so that it is compatible with IT systems. A number of the most common data formats are described in Table 3.5.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spreadsheet</td>
<td>Excel spreadsheet or spreadsheet from ‘Office-like’ software</td>
<td>Little IT experience required in order to manipulate data</td>
<td>Large file sizes; limited flexibility with data storage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Can analyse data straight away</td>
<td>Does not hold data in multiple relations like Access or SQL</td>
</tr>
<tr>
<td>Text files</td>
<td>A string of text delimited by comma, tab, semi-colon etc.</td>
<td>Easy to import to a database, for example Excel or SQL</td>
<td>Not able to immediately analyse data in this format</td>
</tr>
<tr>
<td>CSV</td>
<td>Stores tabular data (numbers and text) in plain-text form</td>
<td>Simple file format, widely used</td>
<td>Each line only corresponds to one record or case in a dataset i.e. ‘flat’ structure.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Most programmes support CSV import</td>
<td>Very simple, but inefficient and inappropriate for datasets with any complex structure</td>
</tr>
<tr>
<td>XML</td>
<td>XML is a mark-up language with tags to define data</td>
<td>XML was created to readably structure, store and transport information for database imports</td>
<td>XML is designed to carry data, not to display data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XML is the most common tool for data transmission for many types of application, especially the web, where it can be automatically compressed</td>
<td>XML syntax is verbose and relatively long</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Structure of XML allows data to be imported into multiple relations in a relational database</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Support of the Web Services technology</td>
<td></td>
</tr>
<tr>
<td>JSON</td>
<td>JSON is a simple but complete data format for web services</td>
<td>Increasingly common, strongly structured, flexible, simpler to use than XML</td>
<td>Not yet widely used</td>
</tr>
</tbody>
</table>
**EXAMPLE 3.24** Health Protection Agency Syndromic Surveillance Service

Country: UK
Data source: telehealth, general practitioner, emergency department
Status: active (1999)

Four syndromic surveillance systems are coordinated by the HPA. Each system is managed separately, with some differences in the time schedules:

- **HPA/NHS Direct**: daily data, analysis and interpretation; weekly reporting
- **HPA/QSurveillance**: weekly data, analysis, interpretation and reporting
- **EDSSS**: daily data, analysis and interpretation; weekly reporting
- **GP Out of Hours**: daily data, analysis and interpretation; weekly reporting

A single summary report is emailed each week to a distribution list of public health officials. This captures the high-level summary messages from each system’s bulletin and provides a link to each bulletin, which are hosted on a website for public viewing.


1. On 1 April 2013 the Health Protection Agency transferred its functions to Public Health England and the HPA/QSurveillance GP system was changed to the Public Health England GP surveillance system.

---

**Transmission process**

Transmission should be electronic, secure and encrypted. There are several mechanisms by which data can be transferred and you may need advice from IT experts:

- Secure email
- Secure web server
- Secure file transfer protocol (FTP)
- Using existing secure reporting mechanisms

**Frequency of transmission**

The more frequent the better:

- Daily if available – even if reporting of surveillance outputs is weekly, it allows flexibility for intra-weekly reporting in the event of a public health incident.
- Timing of transmission – early in the day if possible to allow for analysis/interpretation for real-time public health reaction.

Data transmission that is more frequent and earlier also allows for errors to be detected and corrected earlier, ultimately resulting in higher data quality.
Centralisation of data transmission

When the syndromic surveillance system is national, there are two possibilities for transmitting data:

- From the data provider directly to the syndromic surveillance system
- From the data provider to a local or regional entity, which then sends the data to the syndromic surveillance system – ‘centralisation’

One of the advantages of centralisation is that it allows for better management of the contact points: it is easier to have one contact point in each region rather than one at each data provider and liaison with data providers is easier because this is done by regional entities, which are in close proximity to the data providers. Another advantage is that it is easier to implement and maintain IT, so that data flow can be managed in a timely and consistent manner.

EXAMPLE 3.25 The French syndromic surveillance system SurSaUD®
Country: France
Data source: Emergency department attendances / General practitioners associations SOS Médecins

Data from the emergency departments and from general practitioners’ associations are transmitted daily using encrypted XML format file by FTP.

Although some of the emergency departments transmit data directly to the InVS server, most of them first centralise the data at a regional level.

Data from SOS Médecins local associations are centralized in a national SOS France central server before daily transmission to the InVS server.

Transmission process from emergency departments and SOS Médecins to the InVS
3.5 How should data be aggregated?

In most cases, syndromic surveillance systems are based upon the collection of anonymised individual records (which are invariably reported in an aggregated form). In the context of data transmission, it is important to consider at which stage this aggregation takes place and how much control the host organisation responsible for the syndromic surveillance system has over it.

Some syndromic surveillance systems use individual records with personal and diagnostic data which are collected in a central repository that is normally maintained by the data provider. This allows for very flexible reporting with new syndromes defined as needed. Since such a central repository contains highly sensitive information, access must be controlled tightly. Such systems are generally expensive. However, national ‘eHealth’ initiatives, which are being increasingly developed, will eventually make such collection more feasible. An example is the Finnish AvoHilmo system. This is based on a central eHealth repository, to which health care providers are obligated by law to send data. The system was developed for purposes other than syndromic surveillance, yet it will serve as a basis for surveillance.

Other syndromic surveillance systems only collect aggregated data. This is more secure, in that the database does not include sensitive data. However, it is not very flexible because it does not allow the definitions of syndromes to be changed rapidly.

It is also possible to combine elements of both types of systems. In this way, some of the flexibility of the centralised system can be maintained, without losing much of the security of the distributed system. A good example is Google Trends, which allows for a relatively free definition of search terms, but the results are transformed to maintain confidentiality.

3.5.1 Aggregation by signs, symptoms and proxy measures

A ‘syndromic indicator’ or syndrome is defined as a group of symptoms or signs or a proxy measure in a patient or population. A syndrome raises the suspicion of the presence of a disease or condition before it is confirmed.

Syndromes as defined above, may contain symptoms specific to a particular condition or, as in most cases, be a group of non-specific symptoms (see Table 3.6). The presence of sensitive/specific symptoms will result in a sensitive/specific definition. A ‘sensitive’ definition is able to identify precisely the presence of the selected syndrome (minimising the number of false negatives), whereas a ‘specific’ definition is able to identify the absence of the selected syndrome (minimising the number of false positives).

The choice between a sensitive or a specific definition depends on the severity of the syndrome and thus on the choice to minimise false negatives or false positives. If the syndrome is very serious or
<table>
<thead>
<tr>
<th>Epidemiological scenario</th>
<th>Syndromic indicators useful for monitoring scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Seasonal/epidemic</strong></td>
<td></td>
</tr>
<tr>
<td>Influenza</td>
<td>Fever, Influenza like-illness</td>
</tr>
<tr>
<td></td>
<td><strong>Respiratory</strong></td>
</tr>
<tr>
<td></td>
<td>Acute respiratory infection (upper and lower)</td>
</tr>
<tr>
<td></td>
<td>Fever, Bronchilolitis, Bronchitis, Cough, Dyspnea,</td>
</tr>
<tr>
<td></td>
<td>Pneumonia, Sinusitis, Sore throat</td>
</tr>
<tr>
<td></td>
<td><strong>Gastro-intestinal syndrome</strong></td>
</tr>
<tr>
<td></td>
<td>Diarrhoea, Nausea, Vomiting, Abdominal pain,</td>
</tr>
<tr>
<td></td>
<td>Dehydration</td>
</tr>
<tr>
<td></td>
<td><strong>Extreme temperatures</strong></td>
</tr>
<tr>
<td>Heat</td>
<td>Dehydration, Heat stroke, Hyponatremia, Malaise,</td>
</tr>
<tr>
<td></td>
<td>Injury (falls), Mortality</td>
</tr>
<tr>
<td>Cold</td>
<td>Hypothermia, Frostbite, Cardiovascular symptoms,</td>
</tr>
<tr>
<td></td>
<td>Respiratory symptoms, Injury (falls), Mortality</td>
</tr>
<tr>
<td>Mass gathering</td>
<td>Respiratory syndromes</td>
</tr>
<tr>
<td></td>
<td>Gastroenteritis (diarrhoea, vomiting)</td>
</tr>
<tr>
<td></td>
<td>Bloody diarrhoea, Botulism-like (double vision),</td>
</tr>
<tr>
<td></td>
<td>Rash, Meningitis</td>
</tr>
</tbody>
</table>

Based on WHO guidelines: [www.who.int/csr/mass_gatherings/en](http://www.who.int/csr/mass_gatherings/en)
considered to be a public health threat, it is preferable to minimise the number of false negatives, using a sensitive definition. On the other hand, if the syndrome is considered as a minor health problem, a specific definition may be adopted.

Depending on the purpose of the system, the same syndrome may have both sensitive and specific definitions. For example, the sensitive definition of gastrointestinal syndrome is based on the presence of five sensitive conditions (diarrhoea, nausea, vomiting, dehydration and abdominal pain); the specific definition is based on three specific conditions (diarrhoea, vomiting and abdominal pain).

A commonly used list of syndromes captured by syndromic surveillance systems can be found in The Communicable Disease Alert and Response for Mass Gatherings: Key Considerations, WHO June 2008 or Chapman et al., 2010.

### Olympic Lessons

It is important to discuss the enhanced surveillance requirements with the Olympic surveillance coordination teams. These include the required indicators relevant for mass gatherings, for example:

- Infectious diseases, both endemic and non-endemic (imported from overseas)
- Bioterrorist-related indicators
- Environmental-related indicators, for example heat and/or cold (if applicable)

<table>
<thead>
<tr>
<th>Epidemiological scenario</th>
<th>Syndromic indicators useful for monitoring scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disasters, natural or deliberate release</td>
<td></td>
</tr>
</tbody>
</table>
| Accidental release of toxic compounds (for example industrial/chemical incident) | Respiratory syndromes (upper and lower)  
Conjunctivitis  
Cutaneous lesions  
Difficulty breathing |
| Natural release of toxic compounds (for example volcano/ash cloud) | Respiratory syndromes (upper and lower)  
Conjunctivitis  
Cutaneous lesions  
Difficulty breathing |
| Bioterrorism                                   | Rash  
Botulism-like (double vision)  
Difficulty breathing |
| **Table 3.6 Examples of syndromic indicators**  |

Infectious diseases, both endemic and non-endemic (imported from overseas)  
Bioterrorist-related indicators  
Environmental-related indicators, for example heat and/or cold (if applicable)
3.5.2 Aggregation by geographic level

Depending on the data collected by the system, there are two different options for geographical aggregation:

- Based on health facility or health unit location: the number of individuals is calculated based on where they were captured by the system (Example 3.26). This could be, for example, the health facility, call centre for ambulance dispatch, telephone helpline call centre, pharmacy or the school. This allows the patient’s health condition to be correlated with exposure to an environmental event (for example, a heat wave) or to consider infectious disease propagation. This option is preferable when you have to analyse data in a place or during a period with a large variation in the population such tourism or during a mass gathering.

For animal health this type of aggregation is not common. Veterinary clinics, slaughterhouses and rendering plants are usually not considered as relevant aggregation levels as their location is not systematically linked to the location of the animals.

EXAMPLE 3.26 The Health Protection Agency/QSurveillance GP surveillance system

Country: United Kingdom
Data source: GP consultations
Status: Active (2004)

Geographical aggregation of flu data to the health boundary (primary care trust) level across England (left) and London strategic health authority (right)

Maps taken from the HPA/QSurveillance weekly surveillance bulletin which can be found at: www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/RealtimeSyndromicSurveillance/

1. On 1 April 2013 the Health Protection Agency transferred its functions to Public Health England and the HPA/QSurveillance GP system was changed to the Public Health England GP surveillance system.
Based on individual’s home/animal location: generally the post code (Example 3.27). In animal health this is the most common level of aggregation (animal or herd location).

In both options, different geographical levels of aggregation can be considered for the data analysis: from the lowest level (health facility or herd) to the highest level (national aggregation). Using the lowest level may mean having to analyse low numbers with higher variability, which would decrease the performance of the statistical detection methods. In contrast, aggregation at higher levels limits the chances of detecting local and moderate variations in the indicators.

Olympic Lessons

It is necessary to provide surveillance of all geographical regions involved in an event. For example, during the London 2012 Olympic and Paralympic Games, there was a requirement to report at national and London level and also in cities hosting other Olympic events.

It is also important to investigate the potential to further analyse data and ‘drill down’ to lower geographical levels in the event of an incident. For example, if an excess of fever consultations occurs during a mass gathering, can individual records be analysed by post code to identify any clustering around the location of the events?
3.5.3 Temporal aggregation

The choice of the temporal aggregation depends on the nature of the disease or issue/event being monitored (see Example 3.28). For example, aggregation by week is preferred for the follow-up of the dynamics of seasonal epidemics, whereas the surveillance of the occurrence of emergent diseases or the impact of environmental events (for example, heat waves, floods and allergies) is generally analysed on a daily basis.

For routine surveillance, the most commonly used temporal aggregation is by day or week, which allows changes over a short period of time to be observed. For aggregation by day, you need to take into account the effect of weekends and bank holidays.

Temporal aggregation by month or longer or by specific periods could be considered for the impact assessment after the occurrence of health events.

**EXAMPLE 3.28** Syndromic surveillance system SurSaUD®/Emergency department network OSCOUR®
Country: France
Data source: Emergency department attendances
Status: Active (2004)

The figures below show the number of visits to emergency departments in Paris resulting in a diagnosis of asthma from May to September 2006. They illustrate how the choice of temporal level can influence the prompt detection of events.

**Visits to emergency departments with a diagnosis of asthma aggregated by day, week and month (Paris, May–September 2006)**

**Daily aggregation**

The daily aggregation clearly shows two major peaks, reflecting a major increase in cases in a few hours, that were related to an increase in allergen levels during two days in June 2006.

In contrast the weekly aggregation did not reveal any abnormal variations and the monthly aggregation appeared to show that the indicator to be stable over the period.
3.5.4 Aggregation by demographic/population information

It is important to be able to identify population groups that have an increased risk of infection or greatest public health risk. Information that is broken down by age is of particular importance as it allows at-risk groups of individuals to be identified (for example, the elderly and young children, calves and adult bovine). Moreover, age may provide useful information in cases in which a particular clinical intervention is age-based (for example, seasonal flu vaccination).

When possible, age divisions should be consistent with the expected frequency of the syndromes in a particular age group (for example, the analysis of bronchiolitis is only of interest for children aged under two years). For comparability purposes, it could be useful to standardise age groups for specific syndromes based on associated infections (there are age groups recommended when performing analysis of EU notification laboratory data).

Although there are no standardised age groups for syndromic surveillance systems across Europe, they do exist in individual countries and are used by other types of surveillance systems. For animal health, it is possible to use age group definitions based on government regulations. For example, European Regulation R361/2008 defines specific age categories for slaughtered bovine (i.e., less than 8 months and 8–12 months) and all European slaughterhouses have to record this information. However, depending on the objective of the system, using a government regulation to define age categories might not always be relevant, although it could be interesting for European comparability. Given that there is no consensus on age category definition, each system has its own.

For human health, in early detection and routine follow-up, aggregation by gender may not provide useful supplementary information, since most diseases affect men and women equally. Analysis by gender, when combined with other data, may complement retrospective impact assessment. For animal health, as explained in Section 3.2.2, aggregation by sex is important because it has an impact on breeding practices and diseases.

3.6 How can I ensure continuous access to data

One of the key attributes of a successful syndromic surveillance system is longevity (i.e., the continued provision of data over a long period of time). To ensure longevity, several issues should be considered. In particular, data providers should not have to undertake any additional tasks to participate in the system, and they should be able to continue providing the data during an emergency (for example, a flu pandemic).

Assessing the level of experience of data providers can provide an indication of the stability of data collection (for example, an established system of data collection is more likely to be stable and undergo fewer changes that may affect the syndromic surveillance system).
3.6.1 Data provider relationships

One of the most effective means of ensuring continuous data provision is to build positive relationships with providers. To this end, it is important that the data providers are able to see the results of the analysis of their data, and its use for public health surveillance and decision making. Data providers should also benefit from their participation in the system. For example, their input should be acknowledged and the results of the analysis of the data, including their scientific interpretation, should be promoted in joint publications. For further information see Section 5 on communication.

In most cases, the host organisation contacts the data providers. Opportunities for contact may come about through:

- Meetings of public health professionals
- Scientific literature
- Existing work/projects with the data provider
- Existing contacts
- Identifying possible ventures through research/knowledge of health systems.

Though less common, data providers may initiate contacts because they see the potential of their data and, by participating in a syndromic surveillance system, they see a way to be recognised as part of a public health network. In this case it is important to encourage the initiators and to collaborate to develop an idea.

It is necessary to identify a contact person or persons within the data provider organisation. They will play a critical role in maintaining a positive working relationship and so ensure the success of the system. The key criteria used to identify a contact person are:

- Someone who has an interest in public health surveillance, is familiar with the principles of syndromic surveillance and values the contribution that it can make
- Someone who is willing to act as a leader supporting the syndromic surveillance system from within the data provider organisation, and as the professional link with data providers (for example, GPs)
- Someone who has a senior position within the organisation and so is able to institute changes if necessary.
Contracts and data sharing agreements

Triple-S strongly recommends having written agreements for data sharing, transfer, maintenance etc. between data providers and the host organisation. However, establishing these contracts can depend on:

- The culture within the data provider and/or host organisation and whether having these agreements is a legal requirement
- Whether the data providers are paid to send data to the syndromic surveillance system. In this case, there will be a legal requirement to have both contracts and data sharing agreements in place.

Contracts and/or data sharing agreements can be established at a national or local level and with a representative body or individual partners. A syndromic surveillance system can include different types of contracts or data sharing agreements. These can be simple documents that include an outline of the background of the work, the data items to be transferred, who will be handling the data and who the senior custodians are from each party. A sample template data sharing agreement is provided in the Appendix.

Providing feedback to data providers

As previously mentioned, giving feedback to data providers is crucial for ensuring data quality (see Example 3.29). This can be provided in a number of ways, including:

- Steering group meetings

Establishment of a steering group is especially important during the development phase of the system, when communication among the organisations is crucial for developing data specifications and methods of data transfer.

Regular meetings should be held covering all aspects of the syndromic surveillance system, including the maintenance of data provision.

These meetings can highlight the benefits of a system and provide examples of case studies where the system has provided public health benefits. The data providers also find this positive feedback useful for promoting the system within their own organisations.

They are also an opportunity to discuss aspects of the project that have not gone well or need improvement. They also allow both the host organisation and data providers to discuss concerns.

- Routine surveillance reports/bulletins

During the development of the syndromic surveillance system, and once the system has been launched, the provision of regular reports and surveillance bulletins illustrating how the data are used for public health surveillance will provide the data provider with positive feedback and represent an output that can be advertised on public-facing websites. Some examples of reports and bulletins are available in Section 5 on communication.
EXAMPLE 3.29 The HPA¹/NHS Direct Syndromic Surveillance System
Country: England and Wales
Data source: Telehealth line
Status: Active (1999)

The HPA/NHS Direct telehealth syndromic surveillance system is managed by a joint NHS Direct and Health Protection Agency Steering Group. This group meets every three months to discuss the progress of the system, identify problems or future issues that may affect it, develop ideas for enhancing the work and discuss any research projects that might have been generated.

A single senior NHS Direct contact is always consulted regarding any issues or problems with the syndromic surveillance system.

Any peer reviewed publications or research projects that use the syndromic surveillance data from the system are always co-authored by NHS Direct and HPA.

This positive collaboration has resulted in a successful syndromic surveillance system running for over 13 years; during this time it has been used to respond to a number of public health incidents and has produced over 16 peer reviewed publications (for example Smith, 2010; Loveridge, 2010).

Further information: www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/RealtimeSyndromicSurveillance/

1. On 1 April 2013 the Health Protection Agency transferred its functions to Public Health England.

3.6.2 Stability of a data source

It is essential to assess the future stability of a data source. This can depend on whether it was developed for a specific event or is a continuing data source. There is little point in spending resources on developing a syndromic surveillance system if the data source is due to be discontinued in the near future.

Resilience of technical infrastructure

It is important that syndromic surveillance systems provide a continuous, fully functional service. System unavailability can be the result of a routine process or event (for example, maintenance or upgrade) or an unscheduled system failure.
Key characteristics in the design of a robust system are ‘redundancy’ and ‘resilience’. For the former, where possible, servers should be mirrored and alternative transmission channels set up (use different providers for this purpose if feasible).

Ensuring resilience is a more complex issue. A system is considered to be resilient if it is able to operate in a degraded state and return quickly to its ‘normal’ state. Resilience is the result of monitoring the system, ensuring its protection, detecting failures and managing system restoration. All backups (data, programs, parameters, system configuration) must be done on a regular basis.

A critical element of the system’s availability is monitoring. This is the key process for preventing system disruption. The system load, quality of transmission, and system warnings concerning the network and processes must be supervised to anticipate possible failure or quickly fix anomalies.

The last element of your system availability is its maintenance, repair and operation (MRO). You must arrange for technical support, which ensures both the maintenance and repair of the hardware and software. Finally, do not forget or minimise the maintenance of your human network: you must regularly meet and exchange information with your stakeholders and data sources. The primary data availability is the availability of your sources!

---

**Olympic Lessons**

Large mass gatherings, for example the Olympic and Paralympic Games, are global events with extensive media coverage and exposure. It is often therefore ‘attractive’ for data providers to contribute to and to be seen as contributing to the public health goals of mass gathering surveillance. This advantage can be used when discussing enhanced requirements of the syndromic surveillance service.

It is also important to emphasise to data providers the importance of stability of data sources during the event, when any problems resulting in a deficient syndromic service would be seen as attracting negative publicity.

Likewise, following the mass gathering it is important to fully recognize the contribution of key contacts at each data provider, ideally from senior management of the host organization, to recognize the contribution of these individuals and/or organizations to the mass gathering surveillance.
There are several ways to minimise threats to the integrity of the syndromic surveillance system, ensuring confidentiality, security and credibility. This section provides an overview of ways to mitigate these threats. More detailed information can be found on national specialist websites, such as:

- The National Institute of Standards and Technology – Computer Security Resource Centre (http://crsc.nist.gov)

3.7.1 Confidentiality

Although the variables collected for syndromic surveillance generally do not allow patients to be identified, data providers may collect certain variables that could be used to directly or indirectly identify patients, including:

- National health care identification number/national identification number of the animal
- Date of birth
- Address and/or full post code

Given that these variables are not always necessary for syndromic surveillance, the possibility of removing these variables from the dataset should be discussed with the data provider.

Data privacy and protection are legal requirements defined in Europe by Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995. This Directive has been endorsed by the EU member states, and has been incorporated into the national legislation of individual countries. It is critical to contact the local/regional/national ethical and governance committees for approval of the data protection procedures adopted in your syndromic surveillance system.

Article 29 working party of the European Commission is the European advisory structure concerning the protection of personal data (http://ec.europa.eu/justice/data-protection/index_en.htm).

It is vital to remember that even if the dataset is ‘non-identifiable’ it should be handled as if it did contain confidential patient identifiable data. The personnel who handle and analyse the data should be trained accordingly and sufficient procedures should be put into place to ensure data safety (see 3.7.2. Security for further information).

Confidentiality entails protecting against information disclosure, and must be enforced during data transmission, storage and processing. Assuring the confidentiality of a syndromic surveillance system is critical. Even if you use anonymous data, you must seriously consider
the confidentiality of your system because the data or the results could be sensitive under certain conditions. Breaches in confidentiality (for example, loss of data, placing semi-identifiable data in a public domain and passing data to third-party users without permission) can seriously compromise the surveillance system, leading to a lack of trust on the part of data providers and ultimately result in legal prosecution in certain circumstances.

It is crucial to obtain agreements between data provider(s) and the host organisation (i.e., data sharing agreements). These agreements should outline the aims of the syndromic surveillance system, the duties of each organisation and the dataset (including all variables) that are to be transferred. The agreement should also outline the security measures for data transmission and storage.

If the data provider collects and extracts data from a number of individual sites (for example, GPs), you should also consider having an agreement in place between the data source and the individual sites, to gain consent for the use of data in the syndromic surveillance system. This also provides individual providers with the option to opt out of the system.

Your local/national ethics committee may require these agreements for the legal acceptance of your system.

### 3.7.2 Security

Data transfer processes need to have high levels of security. This is normally achieved through encryption to make it unreadable to an unauthorised party.

There are different levels of encryption:

- Data Encryption Standard (DES) (64 or 128 bits);
- Triple-DES, a more secure version of DES (168 bits);
- Advanced Encryption Standard (AES) (128 or 256 bits).

It is also possible to consider applying encryption to the surveillance data only, or to use a fully encrypted transmission system.

In the first instance, simple systems can be employed, such transmission of data as files attached to emails or by File Transfer Protocol (FTP). Fully encrypted systems involve merging the cryptographic and transmission infrastructures; in these instances, Secure File Transfer Protocol (SFTP) or the Hypertext Transfer Protocol Secured (HTTPS) are usually used.

As a compromise between these two solutions, Virtual Private Networks (VPN) can be used. This technology creates a private transmission infrastructure in which it is possible to use any means of transfer in a protected mode.
Security of data storage

The storage of surveillance data is of highest importance, in that these data must be safeguarded against intrusions and unauthorised access. Servers storing data must be protected behind firewalls and secured from viruses or worms.

Surveillance reports resulting from the surveillance process must also be protected, given the potential for these reports to contain sensitive (even if not identifiable) information.

Surveillance data is often stored and transported on external data storage devices (for example, copies, temporary data storage on laptop computers) and host organisations generally have strict security policies for protecting these data. However, if there are no procedures in place it is important to adhere to good working practice.

3.7.3 Accountability and data quality

Whereas the reliability of the overall system is a natural consequence of the above requirements being met, the credibility of your system and its results depend on the information the system conveys and displays. Two main concepts drive this requirement: accountability and quality.

Data must be sufficiently accurate and valid to fit their intended use. However, the inherent nature of syndromic surveillance systems is that the data are pre-diagnostic or non-diagnostic information (thus not confirmed diagnostic or validated), and it is vital that both the organising institution/organisation and the data providers are aware of these limitations.

An even more important aspect of data quality is its stability. As stated above, unspecific data with possible mistakes in terms of diagnostics are an inherent feature of syndromic surveillance; however, as long as the quality of the data is stable, changes in trends will still be identified. Problems can come about through fluctuations in data quality, which can have a major effect on the reporting from the syndromic surveillance system. These fluctuations may be caused by:

- Falls in the level of diagnostic coding or presenting symptoms, resulting in lower numbers for the numerator
- No receipt of data from certain regions/areas, resulting in a lack of representativeness
- Individual sites failing to report data
- Turnover of personnel with a lack of training
- Loss of motivation and involvement on the part of data providers.
To ensure the credibility of the syndromic surveillance system, it is critical to set up a data quality assurance system. This can be achieved by developing a data quality log at the host organisation that identifies problems in the dataset. Some examples of data quality checks include:

- The number of records received (within expected levels?)
- The structure of each data field (i.e., text/numeric)
- The percentage completion of each data field (within expected levels?)
- The percentage completion of coding fields (do the cases contain the expected level of coding data?)
- Redundancy checks: include total sums as well as individual records for double checking, utilise overlaps by including 2–7 days data in daily uploads
- Serial numbering of uploads/records to detect missing/incomplete uploads.

Monitoring data quality is essential for the success and stability of a syndromic surveillance system. It is crucial to:

If a problem with data quality is due to a problem or change at the data provider, then they should be given detailed information about the problem. It is also advisable to provide positive feedback on the stability of data quality.

If there are issues with data quality, they should be described in all reports or other outputs of surveillance. For example, if poor data quality has prevented certain regions from being represented, or if the quality of syndromic coding is poor, caveats should be included in the surveillance reports to make the readers aware of this. This will prevent the over-interpretation of data. This process also improves the transparency of the system (i.e., not hiding the negative aspects) and can help the data providers to rectify the problem (they will not want to see negative press regarding their data).

Make sure that discussions around data quality have been held prior to the development of the system. This will ensure that when these issues occur, both the host organisation and the data providers are clear about responsibilities, this will improve the efficiency with which the problems are rectified.
Triple-S recommendations for improving data comparability among countries include:

- preferably choose data sources already used by existing syndromic surveillance systems in European member states
- providers of data to a single centralised data source must use methods and organise and present their data in ways that are as similar as possible (for example, file structures, coding systems, choice of variables)
- for all sources, start by using the following dataset:
  - Demographics: age, species and breed
  - Administrative information: date of visit/call/query, geographical identifier (health facility/call centre/pharmacy and/or patient’s post code/herd location)
  - Syndrome (group of symptoms or proxy measures): symptoms/diagnosis/drug(s)
  - Medical/health information: symptoms, signs, medical diagnosis, chief complaints, drugs or syndrome (group of symptoms or proxy measures)

To analyse data in a homogeneous manner and permit European comparability, syndromic surveillance systems should aggregate individual data in a similar way to build and track comparable groups. It would thus be helpful for systems to use a common protocol to determine the diagnoses, symptoms and signs to define common syndromes. If a diagnosis is not coded with an international codification, it is important to verify that, at least, the definitions are similar among syndromic surveillance systems in different countries.

This common protocol should also include definitions of variables groupings, such as age groups. Different age groups could be considered, such as adults (15 to 75 years), children (less than 15 years), the elderly (75+ years) and the youngest (less than 2 years).
3.9 Checklist

- Decide upon the purpose of the syndromic surveillance system
- Assess the availability of data sources
- Evaluate the usefulness of available data sources
- Identify data sources that have the potential to be used for the syndromic surveillance system
- Contact data providers
- Establish communication with data providers and create a steering group
- Determine a main contact within the data provider organisation
- Evaluate the available variables
- Decide whether to collect a minimum or enhanced dataset
- Discuss the preferred format of the data with information analysts and epidemiologists in the host organisation
- Discuss the format of the data with the data provider
- Discuss the available coding system(s) with the data provider
- Undertake mapping of coding systems to a list of syndromic surveillance indicator(s)
- Determine the required modes of data transmission
- Assess the requirements for the frequency of transmission
- Discuss issues of data security and confidentiality with the data provider
- Draw up and sign any required data sharing agreements or contracts prior to data transmission
- Establish a data quality monitoring system within the syndromic surveillance system
- Obtain a data sample from the data provider and validate the data using existing surveillance systems (for example, syndromic surveillance or other)
The following list is based upon the experiences of Triple-S partners and is designed to provide some key advice with respect to setting up syndromic surveillance systems:

- Use existing data sources and avoid any added burden to data providers
- Privilege the use of electronic data transmission
- Promote the incorporation of automated processes wherever possible
- Use the most timely period of reporting
- Perform data quality assessment and develop follow-up tools
- Do not underestimate the usefulness of networking and the provision of timely feedback to data providers
- Try to integrate public health action into syndromic surveillance (i.e., getting alerts and messages from syndromic surveillance systems to front-line public health officials)
- Ensure that all procedures have been cleared by governance and ethical groups
- If entering into legal contracts, ensure that data specifications are clear and have been checked by legal representatives
- Have a clear set of standard operating procedures (SOPs) defining the processes (both automated and manual) that are required to run the data capture and surveillance task processes.

3.11 References


This section provides recommendations for analysing syndromic surveillance data following a series of steps, represented in the flow chart below. Each step is discussed in detail.

**Figure 4.1 Flow chart of key stages in data analysis**

### Step 1
Prepare data for construction of statistical indicators

1. a Select an appropriate indicator to be monitored
2. b Group and pool data appropriately
3. c Consider factors which will influence the choice of indicator(s) and modify data as appropriate

**DATA CONSTRUCTED FOR STATISTICAL ANALYSIS**

### Step 2
Select and implement appropriate statistical method(s)

2. a Describe data and build a baseline or reference group
2. b Select appropriate statistical methods(s)
2. c Assess performance of method chosen (including sensitivity and specificity)

**IMPLEMENT STATISTICAL METHOD CHOSEN**

### Step 3
Investigate statistical signal

3. a Verify data quality and confirm a statistical alarm
3. b Investigate further using epidemiological analyses
3. c Confirm signal
3. d Perform risk assessment of signal

**MAKE A DECISION ON NEED FOR PUBLIC HEALTH ALERT**

### Step 4
Translate signal into public health alert/action

Decide how and who to communicate public health action

**COMMUNICATE PUBLIC HEALTH ALERT**

The main objective of the data analysis phase in a syndromic surveillance system is to produce outputs that correspond to the specific purposes of the system (detection, early warning, situation awareness).

Two main approaches should be considered for analysing data and selecting the appropriate statistical methods according to the purpose of the system:

**Prospective analysis** is used for the early detection of expected or unexpected events and to provide reassurance that there has been no impact on public health. This is the main means of analysing...
syndromic surveillance data. A prospective analysis generally entails using the automated tools and simple methods, routinely applied to many datasets, in order to be able to analyse a large number of health indicators, age groups and/or geographical levels in a brief period of time. The choice of specific method of analysis depends in part on the type of event to be detected (expected or unexpected, or with acute or progressive impact).

Retrospective analysis is used to quantify the public health impact of an event during or after its occurrence, as well as to evaluate the prospective analysis that has been performed. This analysis allows more complex methods to be implemented, without the constraint of having to provide results in a short period of time. Moreover, a precise impact assessment requires taking into account the effect of internal and external factors that can influence variations in the indicators.

Data analysis can generally be divided into four main steps:

1. Preparation of data for the construction of epidemiological indicators, based on the aggregation of the data collected (for example time unit, geographical level, syndrome groups, age groups etc.).

2. Selection and implementation of an appropriate statistical method(s) to determine a baseline reference or background/expected levels, and to provide criteria to assist epidemiologists in interpreting the data. The choice of method must take into consideration the objectives of the system, the characteristics of the data collected and any factors that may influence the accuracy of the data.

3. Assessment of the epidemiological relevance of the output generated by the statistical analysis. This step consists of transforming the statistical signals generated by the methods into potential public health alerts.

4. Interpretation and translation of results of the data analyses into public health action according to the objectives of the system.

Each of these steps of the data analysis is discussed in more detail.

Olympic Lessons

Planning the analysis of data for a mass gathering must be made well in advance of the event. All analyses must be designed, developed and thoroughly tested beforehand to ensure that systems are functioning without errors and that the team performing the analyses are familiar with the processes involved, including troubleshooting.

It is also advisable to be transparent with Olmypic coordination groups regarding analyses. This will ensure that there is an understanding of how data are analysed and will also give some context when results are interpreted.
4.1 Preparation of data

4.1.1 Selecting appropriate indicators

The term ‘epidemiological indicator’ refers to how the data are expressed, for example, as a proportion, crude frequency, or rate. It is monitored routinely to determine whether there are abnormal changes in the frequency of the given event.

There are three types of indicators to consider:

1. **Absolute indicators** expressed in terms of absolute measures such as the number of cases that are used for assessment and communication of total impact.

2. **Relative indicators** are expressed in terms of excesses or ratios such as percentages or incidence rates. These are generally used for comparisons against a baseline or among several population groups (age groups, geographical levels, pathologies). Indicators such as Z-scores (also known as standard scores) are normalised to remove unimportant features of the data.

3. **Statistical indicators** are outputs from statistical analysis such as test statistics, P-values or alarms. They are summary indicators that reveal a significant difference from a normal situation and quantify the degree of departure from this normal situation. They are useful for inference.

To define the indicators to be monitored, it is necessary to identify the data providers, their number (for example, all of the emergency departments in a region; 10 of the 100 GPs in a city) and when their involvement with the system commenced. This is important for defining the denominator used for calculating frequencies or rates and to take into account the scalability of the system (progressive participation of data providers) (see Example 4.1 and Box 4.1).

**EXAMPLE 4.1 SurSaUD®**

**Country:** France

**Data source:** Emergency departments and general practitioner associations, SOS Médecins

**Status:** Active (2004)

SurSaUD® combines four data sources including emergency departments and general practitioner associations. In 2012, 370 of the 650 emergency departments were in the OSCOUR® network, covering 60% of emergency visits nationwide. All French regions are represented although coverage within the regions varies. In 2012, 59 of the 62 general practitioner associations also contributed to the system, covering the largest urban areas and one overseas territory.

The partial and heterogeneous coverage of the networks has to be considered when interpreting data and comparing results from different regions.

Syndromic surveillance based on data collected by a network with the progressive inclusion of new data providers (for example emergency departments, ambulatory centres) can mean that there is a false increase in the quantity of indicators. The following examples illustrate two solutions to eliminate the effects of the scalability of the system due to the progressive inclusion of new data providers.

**Correction by using percentages**

The entry of each new emergency department, marked by a vertical blue line on the graph, leads to an increase in the number of total attendances. When analysing the crude number, it is difficult to distinguish the contribution that is due to the scalability of the system from that due to the influence of public health events. The use of percentages limits this effect.

**Correction using a constant sample of data providers**

Data aggregation is done by using a constant/stable number of data providers for the period of analysis under consideration. When the OSCOUR network was set up in 2004, 23 emergency departments were involved, this increased to 220 by 2010 and to 350 by 2012. The selection of the sample of emergency departments is based on the dates of the first and last transmissions from each data provider and is automatically updated after each daily data transmission. An automated selection of the data providers has to be implemented considering these dates.
4.1.2 Grouping, regrouping and pooling of data

When data are collected by the data provider, the way in which they are aggregated depends mainly on external constraints; whereas when the data is analysed, the researchers/epidemiologists managing the syndromic surveillance system determine how the data are to be aggregated.

Although aggregation leads to loss of information, it facilitates analysis, interpretation and presentation of the findings. Statistics that refer to smaller aggregated groups provide findings with higher uncertainties, yet a minimum level of detail may be necessary to interpret the data. In determining the unit of aggregation to form groups, you should try to balance sensitivity with specificity. For the surveillance and detection of unexpected events, sensitive groups are preferable. After analysis, the results can be pooled to gain statistical power and strengthen the public health message.

We need to keep in mind that when an unexpected event occurs, health professionals are not always aware of it. Thus at the beginning of the event, the coding used for diagnoses related to the event may be less specific than it would have been if the event had been clearly identified. As the event is identified (becomes public), the coding may become more specific.

When epidemiologists aggregate and analyse the data, they have to place themselves in the same situation as the health professional who does the initial coding. To this end, communication with the health professional can be useful to understand how they code the medical information according to the health situation under surveillance. However, as a general rule, to ensure comparability, the coding practices should not be influenced during the event.

For retrospective analysis, focused groups related to the event can be created that are different from the population groups used for routine surveillance. There is also statistical methodology that can be used to create more meaningful groupings. This applies mostly to syndrome definitions and geographical areas (Chapman, 2010).

4.1.3 Factors influencing indicators

The main challenge in syndromic surveillance is to identify a signal in the data that corresponds to an outbreak or cluster and to follow up the public health impact of the event.

After excluding signals that are caused by miscoding or an artefact, interpretation of temporal or spatial fluctuations of epidemiological indicators (for example, emergency department visits for a specific medical reason or the number of calls to a telephone help line) implies an awareness of the factors that could influence the indicators. Taking into account these factors in a statistical model overcomes their effect on the epidemiological indicators and makes the distinction of a real unusual outbreak easier.
Table 4.1 Factors that influence the indicators

<table>
<thead>
<tr>
<th>Internal factors</th>
<th>External factors influencing the population’s health</th>
<th>External factors influencing health-related behaviour or the organisation of health care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trend</td>
<td>Climatic factors (extreme events)</td>
<td>Vaccination strategy</td>
</tr>
<tr>
<td>Seasonality (months, season)</td>
<td>Environmental (pollens, pollution)</td>
<td>Intensive communication in media</td>
</tr>
<tr>
<td>Day-of-week effect</td>
<td>Current epidemics</td>
<td>Change in the organisation of data providers (merging, closing down)</td>
</tr>
<tr>
<td>Bank holidays</td>
<td>Industrial accidents</td>
<td>Coding change</td>
</tr>
<tr>
<td>Special days (e.g. music festivals, strikes or public unrest)</td>
<td>Mass gatherings</td>
<td>Software modification</td>
</tr>
<tr>
<td>Geographic clustering</td>
<td>Media effects</td>
<td></td>
</tr>
<tr>
<td>Age and gender</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are three main categories of factors (see Table 4.1):

1. **Internal factors**: intrinsic characteristics of the indicators, such as a trend or seasonal factors (see Example 4.2). They may reflect changes in the use of health care by the population. Internal factors can also include factors related to the characteristics of the population under surveillance, such as age or gender. Such factors can be introduced in the model, or the epidemiological indicators can be analysed by age group or gender.

2. **External factors influencing the population’s health**: contribute to fluctuations in the epidemiological indicators. These can include climatic factors related to extreme events (heat waves/cold spells, floods, storms) or other environmental factors such as pollens or pollution and factors related to industrial accidents. See Example 4.3.

**Olympic Lessons**

The population movements associated with mass gatherings can involve:

- Incoming visitors/spectators
- Incoming teams/participants in the event
- Incoming press/VIPs/ambassadors
- Reduced numbers of seasonal tourists avoiding the disruption of the mass gathering
- An exodus of local residents moving away from affected areas to avoid the disruption of the mass gathering.

All of these have the potential to impact on health care usage – affecting numerators, and the local population – affecting denominators. It is therefore important, in advance of the mass gathering, to undertake work to assess the potential impact of population changes on syndromic surveillance systems.
EXAMPLE 4.2 Public Heath England – GP Out-of-Hours Surveillance System
Country: England
Data source: GP Out of Hours
Status: Active

The variation of the epidemiological indicators could be explained by the distribution of the population under surveillance. The day-of-week is an internal factor influencing the indicators.

Daily number of visits to out-of-hours services in UK, November 2009–May 2010
In the Figure it can be seen that additional important peaks are observed at the beginning of the week and during bank holidays.

EXAMPLE 4.3 SurSaUD®/emergency network OSCOUR®
Country: France
Data source: Emergency department attendances
Status: active (2004)

Climate situations can act as external factors that affect the epidemiological indicator under surveillance.

Assessment of the heat indicator, summer 2006
The Figure shows that variations in temperatures during a heat wave are closely correlated to the variation in the heat indicator (malaise, dehydration, hyperthermia and hyponatremia).

Source: Josseran et al., 2009.
3. **External factors influencing the behaviour of the population regarding health or the organisation of health care**: examples include changes in legislation or in vaccination strategies. Changing the codes or the definitions used for reporting symptoms or medical diagnoses may also contribute to fluctuations in the indicators with respect to historical data (see Example 4.4).

**EXAMPLE 4.4 SurSaUD®/emergency network OSCOUR®**

**Country**: France  
**Data source**: Emergency department attendances  
**Status**: active (2004)

**Proportion of medical diagnosis notified by two groups of emergency departments**

The Figure illustrates an external factor influencing behavior in the organisation of health care.

While for the first group (blue) the proportion is stable during the entire period, in the second group (green) a major decrease began to be observed in September 2009 due to a coding change. This change has to be taken into account in selecting the historical data to estimate the baseline.

Source: Banzet et al., 2013.

---

Note: it is useful to keep log of the dates of special events. These events may constitute factors influencing the data for future analyses. Knowledge of their occurrence allows these factors to be taken into account in the historical data and limits their influence on the performance of the statistical model.

The statistical methods generally used in syndromic surveillance are designed to take into account the main internal factors, particularly trends and seasonality. Some more complex methods, such as regression methods requiring statistical software, can take into account both internal and external factors by introducing them as variables in the statistical model.

When it is not possible to control for the influence of some factors in the statistical model, implementation strategies can be proposed to avoid their influence. For example, only part of the historical data can be used to determine the baseline, by selecting the equivalent periods (or days). Finally, their potential influence has to be taken into consideration during the interpretation of the results.
4.2.1 Introduction
The choice of statistical methods depends on the following criteria:

- The objective of the analysis, particularly the type of event: event with acute impact v. event with progressive impact, seasonal or expected events v. unknown or unexpected events
- The main components of the data, such as:
  - internal factors (trend, seasonality)
  - low or high number of observations (Poisson or Gaussian distribution)
  - autocorrelation between successive or adjacent observations
- Availability of historical data
- Coverage in terms of geography and age distribution
- Availability of software and information technology (IT) resources
- Availability of statistical expertise.

The statistical analysis needs to be done in a timely manner, especially if it is a prospective analysis. Numerous indicators to cover various pathologies, age groups or geographical levels should be analysed, which over time will result in numerous analyses being performed. This is why the analysis should be as automatic as possible, yet without losing the ability to make adjustments. This will eventually affect the choice of the software used.

While a variety of methods can be used to perform statistical analyses, the choice may depend on the resources available. The methods can range from simple descriptive analysis (for example, Excel or other spreadsheet programs) to more complex methods requiring such statistical software as R, Stata or SAS. Later in this section we discuss the implementation of statistical methods and IT tools for data analysis.

Although various studies have tried to identify the most appropriate statistical methods for detecting an outbreak according to time series or spatial characteristics, there is no single ‘perfect’ method for automated syndromic surveillance systems. Given that the literature contains numerous studies on the various methods and comparisons of their performance (Jackson et al. 2007, Hutwagner et al., 2005, Fricker et al., 2008), in this document we focus on the main structure of the analyses, with a limited number of examples for each step.

4.2.2 Description of data and building a reference
The visualisation or display of the data fulfils two purposes:

1. It is used to describe the data, which contributes to selecting the most appropriate statistical methods to model the data. The goals of this description are to identify the main components of the data (internal factors, distribution), to control data quality in terms of stability and completeness, and to check the occurrence of aberrations in historical data.
2. Visualisation is also used to *analyse* data as part of routine public health surveillance. In fact, much can be achieved using simple descriptive analyses. A well-designed graphical display can reveal aspects of data that even the most advanced statistical methods cannot. The power of human interpretation should not be underestimated. Although syndromic surveillance must be able to detect several different types of signals, most statistical methods are fine-tuned to detect just one type. Moreover, the outputs of the analyses must be interpreted and communicated to their intended audience, and visualisations are a helpful tool.

The method of graphical display used will vary depending on the type of data being analysed. For time-series analyses, a line graph is most appropriate. For spatial data, maps are useful (Example 4.6), although tables can also be used. Many plots can be produced using normal office tools, such as Excel or database reporting tools. General statistical packages (SAS/Stata/R/SPSS) can produce most kinds of graphs and other illustrative methods with relative ease. Maps typically require GIS (geographic information system) tools, but general statistical software can also be used.

For the description of data, different kinds of distribution plots, such as box-and-whisker (Examples 4.6 and 4.7) and density plots, can complement time-series plots. Such tools contribute to describing the population for different variables.

The message of a graph is actually quite simple: “These are similar and those are different”. However, such comparisons should be clear, evident and fair, as discussed below.

---

**EXAMPLE 4.5 NHS 24**

*Country*: Scotland  
*Data source*: calls to the telehealth line NHS24  
*Status*: Active (2005)

**Proportion of all NHS24 calls for fever – intermediate geographies Scotland**

The map illustrates the geographical distribution of calls for fever as a proportion of the total number of calls in the different counties in Scotland.

This output is useful in that it provides a good visual overview of the health situation for fever in the country.

To be able to interpret the results from the different locations, percentages have been used instead of the crude number of calls.
Clarity can be achieved by limiting the amount of detail contained in the graphs. Simple statistical tools can be used to simplify the message. For example, a seven-day moving average will eliminate most of the variation related to the day of the week (Figure 4.8). The outputs of more detailed statistical methods will provide additional tools for clarification.

EXAMPLE 4.6 SIDARTha
Country: Austria
Data source: Emergency medical dispatch
Status: Active (2012)

Daily distribution of the calls recorded by the emergency medical dispatches in Tyrol, Austria

The Figure illustrates the use of a box-and-whiskers tool to support the description of data and identification of internal factors.

The number of calls markedly decreases on Saturday and Sunday, whereas on weekdays it is stable and comparable.

This tool fulfils the need to take into account the day-of-week effect for interpreting this data.


EXAMPLE 4.7 OMAR
Country: France
Data source: Rendering plants
Status: Pilot phase

Monthly mortality rates by production type and age group, France

The Figure illustrates the use of a box-and-whiskers tool to support the description of data and the identification of internal factors.

The number of deaths and calving rate varies markedly by month and production group.

• Signal or effect can be made **evident** by choosing the appropriate scales. Adding a frame of reference also helps. The simplest means of doing so is to include data from previous seasons or weeks. The output from the statistical analyses also helps. Placing markers on relevant dates on time series or point sources on maps will emphasise the message.

• **Fairness/accuracy** results from the use of relevant indicators: comparing numbers of cases from areas with different sized populations is not fair, whereas comparing incidences or percentages usually is.

Since visualisation is the main tool for statistical analyses in syndromic surveillance, there are a number of examples of it throughout these guidelines.

Ensuring that the health situation is evaluated efficiently, particularly in syndromic surveillance, requires more than performing a simple descriptive analysis of the data. In fact, given that the first objective of such surveillance systems is to detect unexpected events, the monitoring of a large range of specific and non-specific health indicators (syndromes) for different age groups and geographical levels is required to be able to cover various health situations in the population. Moreover, in the framework of a prospective approach, the follow-up of all of these indicators must be done regularly.

Statistical methods applied to the different indicators by syndromes, age groups and geographical levels will help to systematise this surveillance by generating alarms when a significant departure from the normal situation is detected (the normal situation is represented by a baseline).
Baselines: providing a frame of reference

The baseline represents the conditions during the non-outbreak period and corresponds to what we would expect to see if there was nothing unexpected going on. It provides a frame of reference (expected values) that the observed data can be compared to. Most, if not all, statistical methods used for syndromic surveillance depend on some kind of baseline estimation. It therefore plays a central role in all data analyses.

In some instances, the baseline is simple: for indicators with a low number of cases such as botulism, the baseline is no cases. Usually, however, the assessment can be very difficult and requires historical data. It may also involve using comparisons between areas or age groups, particularly if historical data are scarce.

To compare the current situation with the past, the baseline is generally calculated either for a similar period in previous years or using the days leading up to the current period. The first option requires long historical data and allows for comparisons with periods that are comparable in terms of seasonality. The second option is preferred when data referring to a short historical period are available. However, with this option, the occurrence of an outbreak will immediately influence the baseline for the following days and can disrupt the interpretation of the analysis during the days after the outbreak.

A number of issues need to be taken into account when estimating the baseline including:

- **How much historical data are available?** Ideally, there should be a minimum of 2–3 years to be able to estimate the baseline accurately and to have experience with more than one previous event. For rarer events, the requirement increases. However, when shorter historical data are available, adapted methods such as control charts are usually recommended.

- **Are there previous events or outbreaks** that need to be taken into account? Events are usually seen as increases in the data. Failing to omit their effect will inflate the baseline and make detection of future events difficult.

- **Are there any internal factors** (see 4.1.3, Factors influencing the indicators) that should to be taken into account? Seasonal variation and trends may hide events that are more gradual; bank holidays may introduce decreases in data that could obscure smaller peaks.

- **Are there any external factors** (see 4.1.3, Factors influencing the indicators) that could be taken into account? Accuracy of the baseline estimation can be improved using any additional information available. Normalising factors are especially useful, as they tend to clean the data.

Note: Baselines should not be confused with denominators, which are observed quantities that measure the size of the population (or population exposure).

For a short overview of commonly used methods, see section 4.2.3.
EXAMPLE 4.9 Pandemic Influenza Primary care Reporting (PIPeR)
Country: Scotland
Data source: GP surveillance for acute respiratory infections in Scotland
Status: Active during the pandemic period (2009–1010)

Daily consultation rates with smooth trend using LOESS and a 15-day window

LOESS (local regression) is a powerful but simple method for fitting smooth curves to empirical data.

This figure demonstrates the visualisation of a temporal baseline for influenza-like illness and acute respiratory infection in Scotland.

EXAMPLE 4.10 GetWell
Country: Sweden
Data source: web queries
Status: Active (2009)

Using web query data to predict outbreak of seasonal vomiting disease in Sweden

Number of queries for *vomiting* submitted to a medical website, compared with the number of laboratory-verified norovirus samples (with baselines and 99% prediction intervals) and the number of media articles about winter vomiting disease.

The baseline is determined by Serfling methods (Serfling, 1963), which are appropriate for estimating the baseline for expected events, such as epidemics. The methods can take into account trends and seasonality through harmonic function.

Source: Hulth et al., 2010.

Olympic Lessons

It is ideal to have several years of hisitorical data to improve the calculation of baselines, thereby strengthening statistical analyses of the data. If this it not possible, the development of statistical tests based upon short-term historical data series can be adopted.
Alarms and thresholds. To complement the baseline, limits are generally imposed in order to determine the amount of deviation from the baseline that will be tolerated before an alarm is raised. These limits constitute statistical thresholds; when the current variations go beyond these thresholds a statistical signal is triggered.

The limits are generally defined by the baseline plus 2 or 3 standard deviations, which corresponds respectively to a risk of 5% and 1% (see Box 4.2). This means that about 5% and 1% of the values are over the baseline by +2 or +3 standard deviations due to the hazard.

Box 4.2 Decision rules for a control chart method

The time series is the black curve and the baseline is calculated by using previous observations for the historical period.

A statistical aberration is launched according to the limit chosen, among the three usual limits corresponding to an excess of the observed values from the baseline of 1, 2 or 3 standard deviations.

On the graph, two periods present observed values over the limits.

In syndromic surveillance, since statistical methods may be implemented on a large number of epidemiological indicators (for different pathologies, age groups and geographical locations), there is a high probability of observing one or several significant signals daily that are due to random fluctuations.

Given that each statistical signal should be investigated in order to confirm or invalidate the potential threat to the population, rules that help epidemiologists to select those signals that need to be investigated can be proposed:

- a limit of 3 standard deviations could be used to identify only the most important departures from the baseline and reduce the number of statistical signals.

- statistical signals could be investigated only if they are observed on at least two consecutive days (see Example 4.11) or are detected by at least two statistical methods simultaneously.

- signals could be investigated only if detected for several epidemiological indicators simultaneously (for example, for several pathologies or different locations).
There are two different scenarios of measuring the public health impact:

- **Prospective measurement:** is carried out while the event is still occurring, usually with limited information from a single or a few data sources. Typically, the current event data is compared to the pre-event data or to the baseline. Sometimes the regional and temporal limits of the event have to be estimated separately.

- **Retrospective measurement:** is carried out after the event is over and can make use of more complex methods to estimate the reference values. Detailed data analyses in terms of the population groups and geographic levels are performed to thoroughly measure the impact. Interpretation of the results can be complemented by considering other data sources, such as laboratory, hospital and mortality data.

The simplest method is to define the time and area of the event and to aggregate the measure of effect over this. For direct excess measures, the inference (usually confidence intervals) can be calculated from the prediction error of the baseline. Statistical measures and regression can be adapted accordingly.

In prospective measurement, one must generally restrict estimating the total direct excess (absolute or relative). It is also possible to take into account retrospectively (usually via regression) the effect of

---

**EXAMPLE 4.11** Syndromic surveillance of epidemic-prone diseases in response to a influx of migrants from north Africa

Country: Italy
Data source: Immigration centre health services
Status: Active (2011)

**Alerts and alarms triggered by notifications from migration centres, Italy**

This figure demonstrates the distinction between alarm and alert.

An alarm (shown as a triangle) was triggered when the observed values exceeded the threshold.

An alert was considered when two consecutive alarms were launched (indicated by an arrow).

As there were no historical data, a simple baseline was defined by using a 7-day moving average, with a 99% threshold considering a Poisson distribution.

Source: Adapted from Riccardo et al., 2011.

**Measures of effect:** statistical indicators

<table>
<thead>
<tr>
<th>Observation</th>
<th>Expected incidence</th>
<th>Threshold (lower limit of the 99% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory tract disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watery diarrhoea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parasitic skin infection</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Adapted from Riccardo et al., 2011.
possible causes, such as the external factors listed in 4.1.3 (for example, temperature).

The statistical analyses are based on indicators. The results of the analyses can also be presented as (meta) indicators, measures of effect. These can be used in the outputs and visualisations in the same way as the original indicators, or as supplementary material. The actual uses for statistical analyses presented in 4.2.3 will be based on these.

The measures of effect can be expressed as direct excess from the baseline, either as a difference or a ratio for, respectively, absolute or relative indicators (see Example 4.12). If necessary, the prediction error of the baseline can be used as a scale of accuracy. The statistical importance of the excess in comparison with the usual variations of the epidemiological indicators is expressed with P-values.

**EXAMPLE 4.12 OMAR**

**Country:** France  
**Data source:** Rendering plants  
**Status:** Pilot phase

**Retrospective quantification of the impact of bluetongue on cattle mortality in the Meuse region of France**

The quantification of the impact is based on the comparison of observed and expected (baseline) deaths for different production types and age groups. Two indicators are used to measure the impact: excess mortality calculated by the difference between the observed and the expected values, and the ratio of the observed and expected values.

<table>
<thead>
<tr>
<th>Production type</th>
<th>Age group</th>
<th>Population (cattle-days)</th>
<th>O</th>
<th>E*</th>
<th>O-E</th>
<th>O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beef cattle</td>
<td>b7 d</td>
<td>324,769</td>
<td>3,921</td>
<td>2,744 [2,633 : 2,860]</td>
<td>1177</td>
<td>1.43</td>
</tr>
<tr>
<td>Beef cattle</td>
<td>7 days – 1 year</td>
<td>13,577,058</td>
<td>5,179</td>
<td>2,633 [2,351 : 2,755]</td>
<td>528</td>
<td>1.2</td>
</tr>
<tr>
<td>Beef cattle</td>
<td>1 – 2 years</td>
<td>9,380,600</td>
<td>540</td>
<td>475 [427 : 525]</td>
<td>65</td>
<td>1.14</td>
</tr>
<tr>
<td>Beef cattle</td>
<td>2 – 5 years</td>
<td>10,648,021</td>
<td>654</td>
<td>536 [486 : 589]</td>
<td>118</td>
<td>1.22</td>
</tr>
<tr>
<td>Beef cattle</td>
<td>N0 years</td>
<td>8,215,083</td>
<td>771</td>
<td>610 [556 : 665]</td>
<td>161</td>
<td>1.26</td>
</tr>
<tr>
<td>Dairy cattle</td>
<td>b7 days</td>
<td>279,806</td>
<td>4,396</td>
<td>3,698 [3,558 : 3,840]</td>
<td>698</td>
<td>1.19</td>
</tr>
<tr>
<td>Dairy cattle</td>
<td>7 days – 1 year</td>
<td>11,894,318</td>
<td>3,604</td>
<td>2,967 [2,808 : 3,131]</td>
<td>637</td>
<td>1.21</td>
</tr>
<tr>
<td>Dairy cattle</td>
<td>1 – 2 years</td>
<td>10,464,190</td>
<td>632</td>
<td>498 [449 : 549]</td>
<td>134</td>
<td>1.27</td>
</tr>
<tr>
<td>Dairy cattle</td>
<td>2 – 5 years</td>
<td>15,557,071</td>
<td>1,775</td>
<td>1,389 [1,304 : 1,475]</td>
<td>386</td>
<td>1.28</td>
</tr>
<tr>
<td>Dairy cattle</td>
<td>N0 years</td>
<td>7,095,367</td>
<td>1,545</td>
<td>1,123 [1,091 : 1,064]</td>
<td>422</td>
<td>1.58</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>87,196,373</td>
<td>21,017</td>
<td>16,691</td>
<td>4,326</td>
<td>1.26</td>
</tr>
</tbody>
</table>

* with 95% prediction interval

Source: Perrin et al., 2010.
4.2.3 Selecting a method

Within syndromic surveillance systems, both unexpected and expected (or ‘known’) events have to be detected. Unexpected events can be considered as unknown events and/or known situations that have never been experienced in the population under surveillance, for example, epidemics of emerging diseases, major industrial accidents, and natural disasters (floods, storms).

By contrast, expected or known events are events that are likely to occur seasonally, such as common infectious epidemics (flu, bronchiolitis or vomiting diseases) or heat waves/cold spells. Even if their occurrence is not observed every season (particularly environmental events), they are expected to be observed during the season (Barnett et al., 2010).

The statistical methods used for detecting events in syndromic surveillance can be divided into five main categories (Lawson et al., 2005):

- **smoothing methods**: the baseline used with these methods is based on observed values from historical data, which can consist of either recent observations (as done with moving averages, see Examples 4.9 and 4.11) or data recorded in previous years (as done with the historical means method). Methods can also use exponential means, giving more weight to the most recent values.

- **time-series methods**: are intended to remove the main components (for example, the trend and seasonal terms) from the initial series by analysing the correlations between the successive observations. The statistical signals are determined by comparing the residual series with the threshold. The most commonly used methods in this category are Box and Jenkins methods (for example, ARMA, ARIMA).

  This category also includes methods based on the combination of harmonic functions (cosinus/sinus) and is often used for expected events, such as seasonal epidemics (for example, flu, bronchiolitis), the most popular being the Serfling method (see Example 4.10).

- **regression methods**: include inference methods, which entail modelling the data by combining various variables. These methods include the General Linear Model (GLM) and other associated methods. Long historical data are generally necessary to fit the model. The Farrington model is one of the most popular methods for early detection in syndromic surveillance.

- **process control chart methods**: were initially introduced by WA Shewhart in 1931 for quality control of the manufacturing processes in industry, and are regularly used for syndromic surveillance. The results are generally presented using a ‘control chart’ to detect statistically significant variations for a syndrome within a population (see Box 4.2).
These methods can be adopted when a short historical period of data is available, as 10 days to 4 weeks of observed indicators may be sufficient to determine the baseline and the threshold.

- Methods incorporating spatial or space-time information: Most of the above methods can be extended to spatial or spatio-temporal settings, though with varying degrees of difficulty. There are special smoothers for spatial data; spatial autocorrelation can be modelled similarly to time-series, etc. Furthermore, spatial scan statistics can be used to detect localised clusters of cases.

As already mentioned, the selection of a statistical method depends on the historical data available, as well as the components of the data series we have to model, which consist of the following:

- **intrinsic characteristics of the data**: using a basic statistical method on data with marked variations (for example, season, day-of-week effect) results in a lower performance in terms of detection.

- **statistical distribution of the data**: methods that use a Poisson distribution are preferred for baselines with low numbers of cases; otherwise, a normal (or Gaussian) distribution is preferred.

Most of the methods are used for detecting unexpected events, though other methods, such as time-series regression methods, are preferred for detecting known events. For retrospective analysis with excess measures, statistical regression can be used as the regression can take into account the effect of possible causes, such as the external factors (for example, temperature).

In addition to the baseline and the measure of effect, the performance of detection methods can vary depending on what is being detected. This can be divided into three broad categories:

1. Detecting **immediate effect** or **acute effect** using the measure of effect and a detection limit (Examples 4.11 and 4.13). Direct effects can be compared against prediction intervals, Z-scores (and P-values) against pre-set limits optimised for sensitivity and specificity.

2. Detecting **gradual effect** using cumulative measures of effect. Typically, this involves some kind of version of CUSUM (cumulative sum control chart), which accumulates standardised measures of effect (Example 4.14).

3. Modelling methods including **change point estimation or clustering**. These are used to model any change affecting the baseline, either immediate or gradual. Similar methods can be applied to spatial and temporal data (Example 4.15).

The selection of methods can also depend on the nature of events that could possibly occur in the series.
EXAMPLE 4.13 NHS Direct
Country: UK
Data source: Telephone helplines
Status: active (2001)

Daily number of calls for diarrhoea in the East Midlands compared with other regions in UK
This figure demonstrates use of syndromic surveillance in the identification of an acute outbreak in 2008.
An important increase in the number of calls is clearly observed during a short period of time, illustrating the occurrence of an acute event.
Source: Smith et al., 2010.

EXAMPLE 4.14 Aster
Country: France
Data source: French armed forces
Active (2004)
CUSUM incidence rate of dengue fever
The figure shows use of a CUSUM control chart for the surveillance of a Dengue epidemic in French Guyana by the French military syndromic surveillance system, Aster, in comparison with two other data sources (military clinical surveillance and civilian biological surveillance) in 2005/2006. The cumulative incidence rate for the syndromic indicator began to increase sooner than for the other two data sources.
Source: Meynard et al., 2008.

EXAMPLE 4.15 SIDARTHa
Country: Austria
Data source: Pre-hospital emergency physician service
Status: Active (2012)
Identification of a space-time cluster of influenza-like illness
This figure demonstrates space-time cluster detection.
The Poisson Cusum detection method was applied to identify a cluster of influenza like illness cases in the district of Goeppingen in Austria from 30 May to 6 June 2009 (during a non-pandemic period of influenza). SatScan software was used for the implementation of the method.
Source: Rosenkötter et al., 2010.
There are three main measures for the performance of the statistical method (Kleinman et al., 2009):

- **Sensitivity**: is the ability to produce a statistical signal that correlates with a real epidemiological alarm. The positive predicted value also used to measure the performance of the methods.

- **Number of false alarms**: number of signals triggered by the statistical model that are not associated to an actual epidemiological alarm. This indicator is related to the specificity. The most commonly used measure is the Average run length (ARL) when there is no change in the system under surveillance.

- **Timeliness**, defined by the delay time in detection, which should be as small as possible. Delay is defined as the difference between the date of the actual occurrence of the event and the date when the methods trigger the statistical signal.

### EXAMPLE 4.16 SIDARTHa

**Country**: Austria, Belgium and Spain  
**Data sources**: Emergency department, pre-hospital emergency physician and emergency medical dispatch  
**Status**: Active (2012)

**Comparison of the performance of a statistical method applied to different data sources**

This table shows the performance of the Poisson Cusum statistical method when used on different data sources. The assessment is based on the comparison of three statisticial measures: sensitivity, specificity and false detection rate.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Sensitivity*%</th>
<th>Specificity †%</th>
<th>Sensitivity*%</th>
<th>Specificity †%</th>
<th>False detection rate ‡%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weekly</td>
<td>Daily</td>
<td>Weekly</td>
<td>Daily</td>
<td>Weekly</td>
</tr>
<tr>
<td></td>
<td>33.3</td>
<td>62.5</td>
<td>7.5</td>
<td>94.6</td>
<td>50.0</td>
</tr>
<tr>
<td>EP-AT</td>
<td>(8/9)</td>
<td>(8/8)</td>
<td>(8/67)</td>
<td>(56/56)</td>
<td>(0/0)</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>60.0</td>
<td>85.7</td>
<td>17.1</td>
<td>96.2</td>
<td>14.3</td>
</tr>
<tr>
<td>ED-BE</td>
<td>(10/10)</td>
<td>(4/7)</td>
<td>(26/70)</td>
<td>(45/53)</td>
<td>(3/13)</td>
</tr>
<tr>
<td></td>
<td>100.0</td>
<td>57.1</td>
<td>37.1</td>
<td>84.9</td>
<td>23.1</td>
</tr>
<tr>
<td>ED-ES</td>
<td>(6/8)</td>
<td>(8/9)</td>
<td>(30/56)</td>
<td>(63/67)</td>
<td>(1/7)</td>
</tr>
<tr>
<td></td>
<td>75.0</td>
<td>88.9</td>
<td>53.6</td>
<td>94.0</td>
<td>14.3</td>
</tr>
</tbody>
</table>

*Brackets: Number of weeks / days with at least one true-positive Poisson CUSUM signal in syndromic surveillance data divided by number of pandemic weeks / days according to reference data.

†Brackets: Number of true-negative weeks / days (no Poisson CUSUM signal) in syndromic surveillance data divided by number of weeks / days outside the pandemic period according to reference data.

‡Brackets: Number of weeks/days with a false-positive Poisson CUSUM signal in syndromic surveillance data divided by all Poisson CUSUM signals of syndromic surveillance data.

Source: Rosenkötter et al., 2013.
Assessing a statistical method in terms of its ability to detect events in terms of the characteristics of the event (shape, duration, impact) and the objective (detection of expected or unexpected events) can be done in different ways:

- **Using real data** collected by the syndromic surveillance system, that includes actual well-identified outbreak(s). In this case, performance is assessed in terms of the ability of the method to detect the actual outbreaks.

- **Using real data** collected by the syndromic surveillance system and including simulated outbreaks in the real data. Performance is measured by ability to detect the simulated outbreaks.

- **Using simulated data including simulated outbreaks.** Performance is assessed in terms of detecting the simulated outbreaks.

- **Comparing real data** with other data from another source, for instance, health data or proxy data (for example, climatic indicators). Performance is assessed by considering the data from the other source to be a gold standard.

In all four of these methods, the ability of the statistical methods to detect the aberration is measured using the indicators already described in this section. The limits of using real data (whether a real or simulated outbreak has to be detected) lies in the fact that other unidentified events can exist in the data; in this case, although the statistical methods could detect them, such events would be considered as ‘false alarms’. By contrast, using simulated events allows the dynamics and magnitude of the events to be controlled in terms of the impact on the population. The performance of the methods can be assessed for both acute and gradual events.

The balance of sensitivity, reactivity and false alarms has to be considered for the final selection of the method. Tools such as ROC curves can help in the presentation of the results (Example 4.17).

### EXAMPLE 4.17 SurSaUD®

Country: France  
Data Source: General practitioners’ associations, SOS Médecins  
Status: Active (2006)

**ROC curve showing the performance of a method based on indicators of sensitivity and specificity**

A ROC graph plots sensitivity on the y-axis against specificity on the x-axis. It shows the trade-off between true positive and false positive. The points above the diagonal line indicate ‘good’ classification results, those below indicate ‘bad’ results. Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular threshold.

Source: Gault et al., 2009.
The implementation of IT tools and the statistical methods for data analysis depend on the following:

- format of the data collected (for example, text file, CSV, XML file)
- size of the database
- IT and statistical expertise
- IT resources available

Most descriptive analyses can be implemented using simple tools. Excel or other spreadsheet programs are well adapted for text or CSV files and can do a great deal, but it may be difficult to automate to the level needed for robust syndromic surveillance. Moreover, such tools can be limited if a large number of individuals are captured by the surveillance system. SPSS software is another option for simple data analysis.

More robust systems can be built using relational databases such as Oracle. These require more IT expertise but facilitate data management and aggregation due to online analytical processing tools (OLAP) or business intelligence reporting tools (Example 4.18). Note that these tools allow only basic statistical methods to be used and regression or inference methods are less adapted for implementation with such tools. Databases are useful tools for collecting, pre-processing and archiving data even when the actual analysis is done elsewhere.

**EXAMPLE 4.18 Diagnosticat**
Country: Catalonia
Data source: GP surveillance
Status: unknown

Use of OLAP tool in surveillance

The Figure illustrates an example of the use of a web-based online analytical processing (OLAP) tool in syndromic surveillance. Its features include:

1. selection of indicator
2. selection of region
3. time series with previous years as reference
4. age structure by gender
5. table with raw data
6. drill down, which allows going into more detail (weekly data).

For a more complex level of analysis, full statistical software that is capable of undertaking syndromic surveillance analyses is needed, with the most popular being R, Stata and SAS. Users can develop their own algorithms to manage data, implement the statistical methods and illustrate outputs with graphs.

Specific packages have been developed for syndromic surveillance, such as the R-package ‘Surveillance’ (Höhle, 2007), which proposes more usual methods, such as regression methods (for example Farrington) and control charts.

For a more complex level of analysis, full statistical software that is capable of undertaking syndromic surveillance analyses is needed, with the most popular being R, Stata and SAS. Users can develop their own algorithms to manage the data and implement the statistical methods (see Box 4.3 for an example). Outputs for the communication reports, such as bulletins, can also be tailored to any specific need, although much expertise may be needed.

For spatial representations, GIS software is recommended, although R and SAS software can produce simple maps (Bivand et al., 2008). For the detection of spatial clusters, the software SatScan can be used, as can Bayesian tools for spatial analysis.

If sufficient IT expertise is available, specific software adapted for analysing the data source(s) of a system can be developed. Such software can provide a user-friendly interface, allowing the consultation of data in accordance with the specifics of the data source and the objective of the system. The data management and statistical methods for the data analysis proposed in such software can be connected with the software listed previously.

The implementations of the database and the statistical methods have to be as automatised and optimised as possible to ensure:

- minimal human intervention and manual actions for the data integration, data management and the execution of the statistical methods (using automated functions)
- a rapid process between the reception of the data and the outputs, in order to ensure a reactive analysis.

Note: SAS or Stata are licenced software, whereas R is free. For more information see R: A language and environment for Statistical Computing www.R-project.org

**BOX 4.3 Using R software for analysis of salmonella outbreaks using Farrington statistical method**

The Figure shows an example of the framework of the R statistical software: the data analysis and data reporting is done by coding algorithms to manage data, implement the statistical methods and illustrate outputs with graphs.

Specific packages have been developed for syndromic surveillance, such as the R-package ‘Surveillance’ (Höhle, 2007), which proposes more usual methods, such as regression methods (for example Farrington) and control charts.

For a more complex level of analysis, full statistical software that is capable of undertaking syndromic surveillance analyses is needed, with the most popular being R, Stata and SAS. Users can develop their own algorithms to manage the data and implement the statistical methods (see Box 4.3 for an example). Outputs for the communication reports, such as bulletins, can also be tailored to any specific need, although much expertise may be needed.

For spatial representations, GIS software is recommended, although R and SAS software can produce simple maps (Bivand et al., 2008). For the detection of spatial clusters, the software SatScan can be used, as can Bayesian tools for spatial analysis.

If sufficient IT expertise is available, specific software adapted for analysing the data source(s) of a system can be developed. Such software can provide a user-friendly interface, allowing the consultation of data in accordance with the specifics of the data source and the objective of the system. The data management and statistical methods for the data analysis proposed in such software can be connected with the software listed previously.

The implementations of the database and the statistical methods have to be as automatised and optimised as possible to ensure:

- minimal human intervention and manual actions for the data integration, data management and the execution of the statistical methods (using automated functions)
- a rapid process between the reception of the data and the outputs, in order to ensure a reactive analysis.
This third stage is what follows the statistical analysis of the data. The identification of an aberration corresponding to a statistical signal needs to be checked and investigated to confirm the reality of the signal and, if necessary, to launch a public health alert. This fundamental process is divided into three main steps:

- verification of data quality and confirmation of the statistical alarm
- epidemiological verification to generate an alert
- risk assessment, to support the decision to launch a public health alert

Each of these steps is detailed below.

This process may be structured using decision algorithms (Figure 4.19).

It is important to note that the team in charge of the syndromic surveillance does not perform all of the steps. In fact, other units or institutions may perform the risk assessment for the population and the actions required to limit the public health impact.

**EXAMPLE 4.19 Scottish syndromic surveillance system**

**Country:** Scotland  
**Data source:** GPs, NHS Direct, Hotel occupational health services, hospital emergency departments  
**Status:** Active (2005)

**Decision algorithm of a multi-source surveillance system**

The Figure shows the decision algorithm from the statistical analysis to the final decision for a public health alert, used during the G8 summit in Scotland, 2005.

Source: Meyer et al., 2008.
Olympic Lessons

The translation of a statistical signal to a public health alert is one of the most important aspects of reporting during a high profile mass gathering. A balance needs to be achieved between having syndromic surveillance systems that are not too sensitive, thereby producing many signals and alerts (i.e. false positives) which can dilute the impact of genuine alerts, and missing public health incidents. It is also important to standardise the way that alarms are assessed: the same processes should be adopted for each syndromic surveillance system, indicator signal, etc to ensure that there is consistency. It is key that there is public health assessment and consideration of potential alerts and not just production of multiple uninterpreted alarms.

- Ensure that the processes involved in translating signals to alerts is clearly documented and covers all stages, including any risk assessments, details of questions/answers, personnel undertaking the alert process. This will provide a clear audit trail in the event of any decisions/alerts being challenged.

- Make sure that all available and relevant information is interrogated for any alarm.

- Do not allow alarms to dictate further epidemiological analyses. In the event of no alarms it is still imperative to scrutinise routine data to ensure that there are no alerts required.

- Simplify alerts by not including complex statistics and make sure that language is understandable for those managing the multifaceted aspects of such mass gatherings.

### 4.3.1 Data quality verification

Once a signal is raised or an aberration is observed, the first thing to check is whether this is due to an error in data collection or a change in the quantity of data recorded (which could be related to the involvement of new data providers or inversely, to a problem with data transmission). Well-defined data quality reports should be readily available. Furthermore, data providers may have to be contacted directly to exclude the possibility of errors.

The results of these controls should then be communicated to those performing the next step of the assessment.

### 4.3.2 Statistical verification

Confirmation of the statistical signal by statistical investigation

In addition to the normal outputs of the statistical analysis, additional more detailed reports are needed to check the performance of the algorithm that resulted in the signal.

These reports could first focus on the implementation of the method, by:

1. Monitoring the statistical properties of the algorithm in order to check that the model correctly fits the data. To this end, the following indicators can be of use: goodness of fit statistics, graphs of residual values and autocorrelations.
2. Analysing the sensitivity of the parameters of the implemented algorithm in order to check the ability of the method to take into account the components influencing the data: amount of smoothing, inclusion/exclusion of factors influencing the data, either internal or external.

3. Monitoring the changes in outputs among historical versions of the outputs.

Supplementary statistical methods based on different algorithms can also be used to confirm or invalidate the departure of the observed indicator from the expected values (baseline). This is particularly useful if the signal has been generated by a basic statistical method. Implementing more complex algorithms taking into account more finely defined internal or external factors can contribute to identifying the potential causes of the signal and support the epidemiological investigation step. Whereas the implementation of new methods in the statistical investigation step has to be considered in a retrospective impact assessment, such an option in a prospective approach (routine surveillance) could be time consuming, since the proper implementation of a statistical method requires validation of the adequacy of the model to the data, which could delay the public health action.

In a prospective approach, several statistical methods can also be implemented simultaneously on the same indicators. This strategy has been adopted in Denmark (Bioarlam system) to follow the evolution of the number of ambulance dispatches and emergency department visits, using a control board.

4.3.3 Epidemiological verification

Epidemiological investigation to validate (or invalidate) the statistical signal and generate an alert

In case of early detection, the cause of the signal may remain unknown and additional investigations are necessary. The epidemiological investigation of the signal starts with an internal in-depth analysis of the data by drilling down to lower levels of data aggregation (or, alternatively, pooling), analyses using more precise age groups, delimitation of the geographical extent of the outbreak. Other associated syndromes or pathologies, based on the same data sources, may be explored. This allows the nature of the health variation to be determined: is it a geographical and/or a temporal cluster?

The data providers can then be contacted for their medical expertise/feedback regarding the situation. Depending on the data source, thorough investigations with the patients may also be undertaken, such as sampling and laboratory tests or consultation of the patient’s medical file.

Note: Setting-up of several statistical methods on numerous indicators with 5% or 1% thresholds increases the chances of generating a statistical signal. Specific solutions can be adopted to take into account the multiple statistical comparisons, such as the Bonferroni correction, which provides a threshold corrected by the number of statistical tests.
When the syndromic surveillance is part of a larger surveillance system, the information collected from other sources should be investigated. In principle, syndromic surveillance is timelier than traditional surveillance systems, but in traditional surveillance, health care workers and/or laboratories can be directly consulted to validate and/or further document the signal, as well as the results from specific systems. Collaboration with experts in the given pathology or external partners working in other teams or with specific systems can help to interpret the observed variations. Alternative data sources, such as public media, can be screened to identify simultaneous events potentially explaining the signal.

Finally, the environmental and epidemic situation may be evaluated by searching for external factors that could influence the epidemiological indicators.

4.3.4 Risk assessment of the threat towards an alert

After verification and confirmation that others need to be made aware of the increase, the statistical alarm becomes an alert.

The identification of the clinical or biological threat from a syndrome is the only step specific to syndromic surveillance system. The purpose of the risk assessment is to characterise the threat, to estimate its potential impact on public health, and to identify control measures in order to limit the dissemination and/or impact of the event.
The risk assessment includes:

1. Characterisation of the threat, which includes the identification of the exposure (biological or non-biological agent), the confirmation of the occurrence (number of suspected or confirmed cases), the description of the health condition, the description of the persons at risk (person, location, time), the identification of the routes of dissemination, and other characteristics of interest.

After characterisation, a threat represents an alert.

2. Estimation of the impact, which includes forecasts of the population at risk, and of morbidity and mortality indicators.

Once the impact has been evaluated, we should be able to answer the following questions:

a. What is the likelihood of the threat spreading in a region, to other regions, to other countries?

b. How severe is the health condition associated with the syndrome?

3. Control measures, which include the preventive and curative measures to be taken.

Note: The risk assessment step may be partly or totally conducted by individuals who are not part of the team in charge of the syndromic surveillance system. However, the epidemiologists involved in the system may contribute to this step, as they can provide supplementary information for interpreting the signal produced by the system.
Data analysis provides outputs showing the evolution of the epidemiological indicators related to health threats and can indicate the impact or the absence of impact, contributing to the reassurance of stakeholders, particularly decision makers.

When an incident occurs, specific bulletins focusing on the potential diseases related to the event can be developed. The scenarios, including a worst-case scenario, should be updated regularly with additional data collected. Global synthesis with an impact assessment is recommended after the end of the event.

Two examples, which illustrate the benefits of syndromic surveillance, include the Icelandic volcanic ash cloud in 2010 and the Chinese melamine milk in 2009. An overview of the health situation related to these events was possible in a very short timeframe after their identification due to the existing syndromic surveillance systems and the availability of historical data allowing comparisons between the current and the previous situations.

### 4.4.1 How should I communicate the results?

In all situations, results from the data analysis should be presented in clear, understandable, self-explanatory graphs, maps, and tables that are systematically interpreted by comments. The comments must be adapted to the intended audience and contain at least a statement of the data source (number of data providers, coverage in the population, etc.), the general methods used to estimate the baseline and the potential limits of the data source.

Each data analysis report must contain the source of the data and a title describing the contents illustrated in the report, with:

- The epidemiological indicator
- The temporal aggregation (daily, weekly, monthly ...)
- The geographical level(s) of the analysis
- The demographic information of the population represented (age group, gender)
- The period of the analysis

Examples of reporting documents are detailed further in the communication section of the guidelines.
The simultaneous identification of syndromic signals in several regions or countries can trigger an investigation to establish the nature of the suspected threat(s). Require rapid communication is then needed among existing systems to identify converging findings.

To ensure the comparability of findings at the European level, data analysis must use standardised epidemiological indicators that monitor fluctuations and measure the degree of departure from the baseline using the same units, whatever the crude frequency and variability of the recorded indicators may be.

Indeed, frequency measurements provided by different data systems in EU countries vary in accordance with the size of the population recorded by the system and with the syndrome definitions. For example, the number of telephone calls to helplines for vomiting diseases in the UK differ from the number of visits to emergency rooms for gastro-enteritis in France, and from the number of web searches for terms related to similar pathologies. For this reason, the Euromomo project, which monitors mortality in Europe, uses the Z-score indicator, which standardises the crude mortality recorded in the different member states.

Syndromic surveillance systems analyse data for two main objectives: (1) early detection of a health event and monitoring the pertinent indicators to track the event’s impact; (2) retrospective risk assessment of the event’s impact on public health.

Each objective has minimum requirements for member states to ensure comparability of data analysis at the European level.

4.5.1 Early detection of a health event and monitoring the pertinent indicators to track the event’s impact

The protocol for the analysis should include common criteria for comparable identification of the health event. Possible criteria include:

- a similar way of defining a reference or baseline and of taking into account factors that influence the indicators under surveillance, such as trends, seasonality and others

- implementation in all syndromic surveillance systems of the same or comparable statistical methods for detecting unusual variations (methods could differ depending on the size of the available historical data, for example).

- the same statistical threshold needed to trigger an alarm should be used. For example, a departure of 3 standard deviations corresponding to a risk of 1% could be defined to identify a significant signal.

- statistical reporting tools, such as graphs, that present outputs in a harmonised way.
4.5.2 Retrospective assessment of the event’s impact on public health

The protocol for the analysis should define for example:

- a unique common study period for events that take place simultaneously in all EU countries. For events that spread progressively across Europe, a different common study period should be defined, such as the period with a significant departure from the baseline (for example, to estimate the impact of an environmental event such as the cloud of volcanic ash, a single period could be defined between the beginning and the end of this event).

- use of comparable epidemiological indicators, such as excess rates or relative risks, to take account of different population sizes and the variability in data source coverage of populations.

At the very least, each country should provide a clear description of the health situation in the event’s specific location; the characteristics of the data sources and findings from the country’s other specific systems; and the epidemiological interpretation of syndromic surveillance findings.
### 4.6 Checklist for evaluation of data analysis

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are the syndromes specific and sensitive enough</td>
<td>Indicators detailed in 3.3.2 and 3.3.3</td>
</tr>
<tr>
<td>Is the analysis and assessment done in a timely manner</td>
<td>Time from data management to the epidemiological validation of the alarm into an alert</td>
</tr>
<tr>
<td>Is the analysis automated</td>
<td>Personnel needed each week in terms of person hours before the epidemiological assessment</td>
</tr>
<tr>
<td>Identification of routine activities that could be automated</td>
<td></td>
</tr>
<tr>
<td>Are the analyses and assessment logged</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Has the performance of statistical methods for detection of events been assessed?</td>
<td>Indicators detailed in 3.3.2 and 3.3.3</td>
</tr>
<tr>
<td>Are all steps of the assessment implemented:</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Data quality</td>
<td></td>
</tr>
<tr>
<td>Statistical</td>
<td></td>
</tr>
<tr>
<td>Epidemiologic</td>
<td></td>
</tr>
<tr>
<td>How are the results translated to public health message?</td>
<td>External evaluation</td>
</tr>
<tr>
<td>Does the output of the analysis fulfil the minimum requirements for external evaluation</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>
References


Perrin J-B et al. (2012). Assessment of the utility of routinely collected cattle census and disposal data for syndromic surveillance *Preventive Veterinary Medicine* 105(3):244–52.


Communication

5.1 What is communication?

Communication is the act of conveying information for the purpose of creating a shared understanding. The word “communication” comes from the Latin “communis,” meaning “to share,” and includes verbal, non-verbal and electronic means of human interaction. In a surveillance network there are two types of communications:

- The type that is done routinely on a daily or weekly basis through networking, production of bulletins, reports, etc. It includes the validation of internal communication among the system’s stakeholders in case of an alarm and the notification of alerts.

- The type performed during an incident that will alter the routine work of the system (for example outbreak, heat wave, mass gathering, etc.). The non-routine type requires more immediate forms of communication such as emails, phone calls and face-to-face meetings.

This section covers why communication is important in syndromic surveillance of both human and animal health systems, what information a syndromic surveillance system needs to provide and which stakeholders are involved. It will provide examples of the types of information and communication methods needed for the different types of stakeholders, both routinely and in non-routine situations. Finally it will also give ‘tips’ on how to deal with requests from different stakeholders.

Olympic Lessons

Good communication is vital during a mass gathering. It is key to make good links with communications team in advance of the event to:

- Determine the communication lines, for example phone numbers and emails during the event
- Determine what will be communicated, to whom and how
- Understand how and where queries for example press/media should be handled
- Manage expectations of communications teams i.e. make sure that they are aware of the surveillance outputs that will be produced and that they understand what the outputs mean
- Agree which outputs are to be made publically available and when/where
5.1.1 Syndromic surveillance communication

When designing a syndromic surveillance system, communication is a key component that should not be neglected. It is crucial at all stages of planning and implementation, and involves all the project participants in raising the visibility of the work.

Any syndromic surveillance system should be part of the general surveillance framework of the country or organisation. It is important that syndromic surveillance communications integrate with, and complement, existing surveillance system communications. Thus, an important initial step when designing a syndromic surveillance system would be to ensure a good understanding of the different health surveillance systems and structures already available, and to contact the owners of the systems to understand their systems and communication mechanisms. In the event of an incident, the syndromic surveillance system key messages have to complement those produced for the incident and not conflict with other messages. This will avoid unnecessary duplication or confusion and will ensure complementary working.

It is helpful to keep an ongoing record of when the outputs have been used (to reassure, in support of an incident etc.). This ‘record of usefulness’ can be helpful in securing further resources.

5.1.2 Why is two-way communications crucial?

Good communication must start within the team. Transparent internal communication will favour a good work environment and save time. Good communication for syndromic surveillance ensures understanding of the aims and objectives and reduces uncertainties in the format of outputs and interpretations.

Furthermore, good communication via feedback is also a key way to motivate data providers and ensure quality data because providers want to provide appropriate data in order that high quality outputs will be produced by the system. A dialogue with the data providers will also increase the chance of detecting changes in data collection or transfer that might in turn affect the output of the system.

In addition, good communication with key funders and policy makers etc. must be maintained to ensure the sustainability of the system. The sustainability of funding can be a ‘tricky’ aspect for syndromic surveillance, because the concept is still quite new. It will be easier to convince funders if good communications demonstrate the value of the system and if the funders are already familiar with its purpose because of previous proactive communications.

Last, but not least, it is necessary to review communications after the system has supported a major incident. This aspect can easily be neglected, yet it is very important to make time for this and to identify what went well, what went badly and to learn from these lessons in modifying and further developing the system. One approach to this is to always include issues about communication into the wider evaluation of the use of the syndromic surveillance in an incident.
5.2.1 Identifying stakeholders

The identification of the syndromic surveillance stakeholders is the first step in the process of communication about a system. Analysis should take into account the significance of the syndromic surveillance outputs to each stakeholders group from their perspective. It is important to keep in mind that not all stakeholders will necessarily share the same concerns or have unified needs or priorities; this is particularly the case for syndromic surveillance as the aims can be quite varied (for example, ranging from early warning for influenza through to assessing impact of mass gatherings). Therefore, results should be presented in a flexible way to a potentially wide variety of stakeholders to enable reactive response when needed. You should also keep in mind that needs are dynamic and that both stakeholders and their interests might change over time, or in the case of a health incident (for example, an influenza pandemic), and that this need for responsiveness and flexibility should be anticipated.

Four main types of stakeholders can be identified:

- Those who will take public health action
- Those who organise or contribute day to day to the system
- Other surveillance professionals and scientists
- General public.

Once identified, it is important to have an updated list of contacts details (address, email, phone/fax depending on the transmission methods for the syndromic surveillance system outputs) for all stakeholders.

During an incident you will not have time to search for telephone numbers or email addresses; you need to maintain an on-going updated list. If possible a person within the syndromic surveillance team should have responsibility for maintaining the list and to undertake a check of this list once or twice a year.

5.2.2 Stakeholders’ expectations

The best way to organise a relevant communication strategy is to ask each stakeholder about their expectations in terms of communication: content, form, frequency etc. You should carefully consider the issue of frequency of outputs. The better the routine communication, the easier communication will be during an incident. Stakeholders will already know each other and the organisers of the system. They will understand the system, its limitations and the outputs. Also, too infrequent production of outputs may not meet the needs of certain stakeholders, but on the other hand providing outputs too frequently risks some stakeholders not reading at all because of ‘information overload’.
Accessing previous stakeholder feedback on other health surveillance activities, if available, can save time and flag up risks, or unresolved issues that can then be prioritised and managed in relation to the communication methods of the syndromic surveillance system.

In addition, we have to keep in mind that each stakeholder has its own ‘language’ and way of working. Hence it is important to adapt the way of presenting and explaining the information for certain stakeholders. For example, the outputs will not be written in the same way for clinicians, researchers, policy makers and media. Underestimating this issue can create real problems, especially during an incident. For example, if the information for the media is not clear and concise (and agreed with those working in complementary surveillance systems) you could create uncertainty and unnecessary difficulties, and potentially undermine the syndromic surveillance system.

Another issue is how to deal with specific requests from different stakeholders. It is important to think through how the syndromic surveillance system will respond to requests for underlying data from various sources, for example media, academics, etc. It is necessary to have clear governance arrangements for the use of the data and this should be discussed among the owners of the system and the data providers at an early stage of the development of the system.

Furthermore, it should be established beforehand how the data providers will be informed about the outputs from the system, and how will they be consulted about, or contribute to, peer reviewed publications.

5.2.3 What to communicate

Several types of information can be disseminated by a syndromic surveillance system depending on the timing and the target audience. It includes:

- Results of analyses, public health interpretations and recommendations for action
- General information on how the system works
- Information on data quality and on how the data are analysed

For example, regular bulletins outlining the current public health situation (trends of seasonal events or information on an outbreak) can be elaborated for stakeholders. Unless otherwise agreed, the output should always consist of analysed and interpreted results in the most appropriate format (data tables and figures).
5.2.4 Examples of syndromic surveillance communication

Depending on the infrastructure of the system, it may be that there is no designated professional communications person assigned to the syndromic surveillance system. Yet, just because there is no designated communications professional, communications should not be forgotten. When planning the development and implementation of a system you need to identify from the beginning what needs to be communicated, by whom, to whom and how. By identifying the aims of the work, the messages and the recipients from an early stage and throughout the work, you will be able to ensure that you collect appropriate data to be able to give relevant information, better adapt the information to be disseminated and make the message(s) clear and ensure effective communication.

Tables 5.1 to 5.4 give examples for the main stakeholders of a syndromic surveillance system and why they need the information, what information they may need, how the outputs could be provided and when. Non-routine activities are usually performed in addition to routine information.

Note: these tables are providing a framework and examples, and thus could be used as a prompt/checklist, rather than attempting to be all encompassing.
### Table 5.1a Communication with the people who take public health action

**POLICY MAKERS**

<table>
<thead>
<tr>
<th>TYPE</th>
<th>WHY</th>
<th>WHAT</th>
<th>HOW</th>
<th>WHEN</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Routine    | Guide health policy decisions for preventing disease, prolonging life or promote health | Graphical presentation of results and written interpretation of results  
Key findings  
Interpretations of detected alerts  
In veterinary field: economic impact (possible decrease of consumption of meat depending on the event) | Email  
Formal meetings  
Designated website | As regularly as possible (for example, weekly) to create a link between policy makers and the syndromic surveillance system | Public Health England routine outputs: www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/RealtimeSyndromicSurveillance/ |
| Non-routine | Guide emergency public health decisions                   | Potential impact (public health, economic, social)                     | Input into incident meetings | Increase frequency, depending on the ability of the data systems (daily if possible) | Daily bulletin for French decision makers during H1N1 pandemic (in French): www.invs.sante.fr/Dossiers-thematiques/Maladies-infectieuses/Maladies-a-prevention-vaccinale/Grippe/Pandemie-A-H1N1-2009-archives/Donnees-de-surveillance |
### Table 5.1b Communication with the people who take public health action

**PUBLIC HEALTH PROFESSIONALS INCLUDING EPIDEMIOLOGISTS CONTRIBUTING TO PUBLIC HEALTH ACTION**

<table>
<thead>
<tr>
<th>TYPE</th>
<th>WHY</th>
<th>WHAT</th>
<th>HOW</th>
<th>WHEN</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Routine    | Inform about the system and its results  | Graphical presentation of results and written interpretation of results | Bulletins, dashboards, graphs sent by email or published on a designated website | As often as the stakeholders want | Routine reporting of Influenza (in Swedish): www.smi.se/publikationer/veckorapporter/webbsok/sasongen-20112012/webbsok-vecka-16-164-224-2012/  
Farmfile (UK) quarterly and annual report: http://vla.defra.gov.uk/reports/rep_surv_cattle.htm |
| Non-routine| Put the alert into context               | Graphical presentation of results and written interpretation of results | Input into incident meetings       | Increase frequency, as the results come out (daily if possible) | Cold Weather Plan by Public Health England syndromic surveillance systems: www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/RealtimeSyndromicSurveillance/SevereColdWeatherSyndromicSurveillance/ |
Table 5.2a Communication with those who organise or contribute to the system day-to-day

<table>
<thead>
<tr>
<th>DATA PROVIDERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYPE</td>
</tr>
<tr>
<td>Routine</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Non-routine</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
### Table 5.2b Communication with those who organise or contribute to the system day-to-day

**STEERING COMMITTEE/ADVISORY BOARD**

<table>
<thead>
<tr>
<th>TYPE</th>
<th>WHY</th>
<th>WHAT</th>
<th>HOW</th>
<th>WHEN</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine</td>
<td>Steer the development</td>
<td>Routine bulletins</td>
<td>Annual report</td>
<td>At least once a year during an annual</td>
<td>Booklet on a French multisource syndromic surveillance system:</td>
</tr>
<tr>
<td></td>
<td>Solve problems</td>
<td>Report system activity, focussing on its</td>
<td>Informative booklet</td>
<td>meeting but can be more regular, for</td>
<td><a href="http://www.invs.sante.fr/pmb/invs/(id)/PMB_9473">www.invs.sante.fr/pmb/invs/(id)/PMB_9473</a></td>
</tr>
<tr>
<td></td>
<td>Share information</td>
<td>public health utility</td>
<td>Oral presentations</td>
<td>example, quarterly meetings</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Meetings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-</td>
<td>Inform about the situation</td>
<td>Written report</td>
<td>Email</td>
<td>Depending on the composition of the group</td>
<td></td>
</tr>
<tr>
<td>routine</td>
<td></td>
<td></td>
<td>Telephone</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Table 5.3a Communication with the other surveillance professionals and scientists

**OTHER HEALTH SURVEILLANCE PROFESSIONALS**

<table>
<thead>
<tr>
<th>TYPE</th>
<th>WHY</th>
<th>WHAT</th>
<th>HOW</th>
<th>WHEN</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Routine    | Inform health professionals and health organisations about the system | Report system activity, focussing particularly on the public health utility i.e. how it has been used and how the data from health professionals have been important | Inclusion of syndromic surveillance information in other reports (for example, influenza, norovirus) | Depending on the frequency of transmission of other reports | Report on influenza where syndromic surveillance information is incorporated into classical surveillance outputs: [www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1287147913271](http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1287147913271)  
Announcing an alert for measles from a syndromic surveillance and traditional surveillance together in a French medical magazine (in French): [www.lequotidiendumedecin.fr/information/infections-enterovirus-un-deuxieme-pic-annonce](http://www.lequotidiendumedecin.fr/information/infections-enterovirus-un-deuxieme-pic-annonce) |
|            | Contribute to the interpretation of output                           | Analysis of outputs                                                   | Lectures/personal feedback from the system owners about the system and utility | During the system set up and when new functions are introduced | |
|            |                                                                      | Alerts and potential public health impact                             | Interview                                                             | Annually to present results of the syndromic surveillance system | |
|            |                                                                      | Feedback on expectations                                              | Article                                                               |                                                                      | |
|            |                                                                      |                                                                      |                                                                      |                                                                      | |
| Non-routine| Verify if there are similar alerts in other surveillance systems     | Outputs of the system and/or alerts                                   | Press release                                                         | As soon as policy makers authorize this kind of communication based on your recommendations and on the existing procedure | |
|            | Inform health professionals on the situation and enhance vigilance   | Recommendations for physicians, medical staff, veterinarians and farmers about the syndromic surveillance | Interview                                                             | Real or near-real-time throughout the incident | |
|            |                                                                      | In the veterinary field: economic impact (possible decrease of consumption of meat depending on the event) | Article                                                               |                                                                      | |
|            |                                                                      |                                                                      |                                                                      |                                                                      | |


Table 5.3b Communication with the other surveillance professionals and scientists

<table>
<thead>
<tr>
<th>TYPE</th>
<th>WHY</th>
<th>WHAT</th>
<th>HOW</th>
<th>WHEN</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine</td>
<td>Foster collaborative syndromic surveillance approach with academics and appropriate acknowledgements of syndromic surveillance and data, rather than simply ‘handing over data’ thus ensuring limitations of the data are understood. Work together to, for example, evaluate the system, research added utility of surveillance, modelling results etc.</td>
<td>May be included in routine outputs. Need to have clear procedure about whether academics can have access to underlying data and approvals needed. May be good to foster collaborations with interested academics.</td>
<td>Website: Outputs usually joint and via peer reviewed publications. Good to encourage peer reviewed publication at all stages of the work so others can learn from the work. Good to foster academic collaborations so the body of scientific evidence for syndromic surveillance can be further built up.</td>
<td></td>
<td>Latest academic collaborations with the Public Health England Real-time Syndromic Surveillance Team (ReSST): <a href="http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/RealtimeSyndromicSurveillance/">www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/RealtimeSyndromicSurveillance/</a></td>
</tr>
<tr>
<td>Non-routine</td>
<td>Expertise for specific studies</td>
<td>Data needed for analyses</td>
<td>As needed</td>
<td>When needed</td>
<td>Elliot et al., 2010</td>
</tr>
</tbody>
</table>
**Table 5.4 Communication with the general public**

**GENERAL PUBLIC (EITHER DIRECTLY OR THROUGH GENERAL MEDIA)**

<table>
<thead>
<tr>
<th>TYPE</th>
<th>WHY</th>
<th>WHAT</th>
<th>HOW</th>
<th>WHEN</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine</td>
<td>Inform the public about the system and its added value</td>
<td>Report system activity, focussing particularly on the public health utility of the system i.e. how it is used</td>
<td>Website Press release Interview Article Social networks*</td>
<td>When the system is set up and for new functions of the system</td>
<td>Announcing an alert from a syndromic surveillance for norovirus (in Swedish): <a href="http://www.smittskyddsinstitutet.se/nyhetsarkiv/2011/vinterkraksjukan-ar-har-/-">www.smittskyddsinstitutet.se/nyhetsarkiv/2011/vinterkraksjukan-ar-har-/-</a></td>
</tr>
<tr>
<td>Non-routine</td>
<td>Inform the population about the situation and assure transparency</td>
<td>It is unlikely that the syndromic surveillance system will produce 'stand alone' press briefings in a crisis/incident but rather the system’s outputs will be incorporated into the overall messages The system’s key messages must complement those produced for the incident and not conflict with other messages</td>
<td>Press release Interview Article</td>
<td>As soon as policy makers authorise this kind of communication based on your recommendations and on existing procedures Will usually be timed to fit in with the overall incident management and thus it is important that the system’s media/communications professionals are aware of its abilities and limitations in advance</td>
<td></td>
</tr>
</tbody>
</table>

* Social media is becoming increasingly used for public health messages. This medium does not differ from other channels.
For purposes of clarity, communication reports from member states should use simple, carefully-worded language; self-explanatory graphs that use scales that make a signal or effect self-evident; the right indicators for the situation being analysed (for example, incidences or percentages if areas with different-sized populations are compared); and standard formats for the outputs.

When an incident occurs that involves a number of countries, syndromic surveillance information should be shared using existing, agreed communication systems and mechanisms already used for other surveillance purposes. Existing public health surveillance networks, for example, ECDC, WHO, EpiSouth, EpiNorth, the EWRS platform, OIE (World Organization of Animal Health), should be used to disseminate syndromic surveillance findings.

When a cross-border incident occurs, it may be useful to share information informally with public health and veterinary professionals working in syndromic surveillance systems in neighbouring countries.

- A syndromic surveillance system should always be an integrated part of the general surveillance framework of a country or organisation and, as such, its output should complement existing surveillance communications (good communications with other surveillance system owners will ensure syndromic surveillance outputs do not ‘wrong foot’ other systems and vice versa).
- Analysed and interpreted results should be communicated – not statistics.
- Information from a syndromic surveillance system should always be clear, concise and agreed with those working in other surveillance systems, to avoid uncertainty and unnecessary difficulties and potentially undermine public health messages.
- Continuous dialogue within the team and between the syndromic surveillance system owners, data providers and different health experts should be ensured at an early stage, when building the system, in order to be able to validate signals/alarms and improve the system and its usability.
- It is helpful to keep an ongoing record of when the outputs have been used (to reassure, in support of an incident etc.). This ‘record of usefulness’ can be helpful in securing further resources.
- Establish clear governance arrangements for the use of the data right from the start of the work.
- When publishing peer reviewed publications data providers should be acknowledged or included in authorship if they have contributed to the publication.
- The list of contact persons/stakeholders should be updated regularly.
5.5 Checklist for evaluation of communication

There are two types of evaluation that should be taken into account: assessment of routine communications and of communications during an incident or unexpected event (non-routine). Both evaluations are important but do not have the same objective.

The following list provides questions to be taken into account during an evaluation of the system:

<table>
<thead>
<tr>
<th>ROUTINE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the communication relevant for each stakeholder?</td>
<td></td>
</tr>
<tr>
<td>Is the frequency of transmission adequate?</td>
<td></td>
</tr>
<tr>
<td>Is the method of transmission secure?</td>
<td></td>
</tr>
<tr>
<td>Is contact lists updated regularly?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AFTER AN INCIDENT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the communication relevant for each stakeholder?</td>
<td></td>
</tr>
<tr>
<td>Was the frequency of transmission adequate?</td>
<td></td>
</tr>
<tr>
<td>Was the information sufficient to help with decision making?</td>
<td></td>
</tr>
<tr>
<td>What other possible measures/actions would have been useful?</td>
<td></td>
</tr>
</tbody>
</table>

For each type of evaluation you may wish to design a dedicated questionnaire and organise a meeting to present the main conclusions (or this may be included as part of a more generic evaluation of the public health surveillance response to an incident). The results of the communications evaluation should be a list of things that went well and things that went wrong. In the case of an incident, this type of evaluation should be conducted immediately after the incident and the results should be made available as soon as possible (no later than the following month), with proposals for improvement (subject, responsible person, deadline).
5.6 References


Evaluation is the systematic collection and analysis of data needed to make decisions both in terms of quality control and improvement. Evaluation of syndromic surveillance systems is optimally done continuously throughout the development of the system. It can be used for orientation during the planning phase and should be carried out at regular intervals to see where the activity stands and where there is room for improvement. Furthermore, evaluation can be used after a public health crisis to identify the added value of syndromic surveillance.

As for any other surveillance system, evaluation of the syndromic surveillance system is important to assess and improve its effectiveness and quality, and for the justification of its existence. The results of the system’s evaluation processes can be used internally and for external communication, for its assessment or for its better understanding and management.

The evaluation can be used to answer five main questions:

- **Are the objectives being met?**
  
  Does the system live up to both the general purposes and the specific purposes defined for it? (see section 1)

- **Is the system functional?**
  
  Referring to all components of a syndromic surveillance system sections 2 to 5

- **Is the system efficient?**
  
  For example, is its operation worth the resources used and output delivered? Does it provide an added value compared with existing systems? Does it produce too many false alerts?

- **Is the system useful?**
  
  If the answer to this question is not ‘yes’ or ‘yes, once fully established’ then the system isn’t worth running.

- **How can the system be improved or optimised?**
Before starting the evaluation it is useful to clarify what the focus of the evaluation is. In the following, three foci are presented that can be evaluated independently of each other or at the same time:

- **The objectives of the system**: measures the extent to which the system has reached the general and specific objectives set for it.

- **The operation of the system**: assesses the system’s functionalities and their planned use. The operation is evaluated without any involvement of its users and with no link to their perception of the system’s benefit.

- **The experience with the system**: evaluates the user’s perspective of the system. It brings valuable information about the way the system is used in practice.

The following section outlines how the evaluation should be carried out, including whom to involve, how the indicators should be measured and selection of the appropriate evaluation study design.

### 6.4.1 Who should be involved in the evaluation?

The evaluation can be accomplished as an internal procedure involving only the personnel performing the activity, but an added value is given if external stakeholders and partners are involved as well. The evaluation can also be carried out by an external party, to get an independent assessment of the system.

### 6.4.2 How should indicators be measured?

The CDC Framework for Evaluating Public Health Surveillance Systems for Early Outbreak Detection (Buehler et al., 2004) is applicable to the evaluation of syndromic surveillance systems and used in practice.

The indicators can be applied to syndromic surveillance systems with all general and specific purposes, not only early detection. Indicators are measured qualitatively and/or quantitatively depending on the kind of indicator.

Table 6.1 lists the indicators from the CDC framework and their use in evaluation of syndromic surveillance systems in terms of: the focus of the evaluation; description of the indicator; measurement of the indicator; and references to studies that have used the indicator for evaluating syndromic surveillance. Additional indicators have been suggested by other authors (for example Meynard et al., 2008).

Though all evaluation indicators are useful, it is not necessary to use all the indicators for each evaluation. The choice of indicators, as well as the specific measurement for each chosen indicator, needs to be defined by the evaluation team and will depend on the system to be evaluated, its specific objectives and the evaluation being performed.
### Table 6.1 Indicators for evaluation of syndromic surveillance systems (CDC framework)

<table>
<thead>
<tr>
<th>Evaluation indicator</th>
<th>Evaluation focus</th>
<th>Indicator description</th>
<th>Indicator measurement</th>
<th>Example of published evaluations using the indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness</td>
<td>Operation</td>
<td>The time taken by the different steps over the course of a health event from the population exposure until the beginning of the intervention.</td>
<td>Timeliness measures the time taken by the different steps, for example, for data transmission and processing including the aberration detection algorithms. It is often measured by comparing the time taken for the same step in another surveillance system. Most often timeliness is calculated for the period between data collection and alert by the syndromic surveillance system. Some specific time to signal measures are used for the algorithm evaluation, such as the different average run lengths (ARL-0, ARL-1) or the average time to first outbreak signal (ATFOS).</td>
<td>Burkom et al., 2004 del Rocio Amezcua et al., 2010 Josseran et al., 2010 Travers et al., 2006</td>
</tr>
<tr>
<td>Validity</td>
<td>Operation</td>
<td>Measurement of how well the system fulfils its objective (for example, detect an outbreak) with the aim to find the right balance between false alerts and detection rate. Validity depends on characteristics of the public health event under surveillance (for example, outbreak) and of the surveillance system (for example, data source).</td>
<td>Measuring validity comprises the calculation of sensitivity, predictive value positive and/or predictive value negative. Validity for a syndromic surveillance system is usually calculated by comparison with a gold standard surveillance system for the same kind of outbreak, for example, sentinel surveillance system for influenza-like illness. If no gold standard exists, validity can be assessed by comparison with another syndromic data source or using the DELPHI method based on a panel of experts.</td>
<td>Centers for Disease Control and Prevention, 2011 Doroshenko et al., 2005 Gault et al., 2009 Guasticchi et al., 2009 Jefferson et al., 2008 Josseran et al., 2010</td>
</tr>
<tr>
<td>Evaluation indicator</td>
<td>Evaluation focus</td>
<td>Indicator description</td>
<td>Indicator measurement</td>
<td>Example of published evaluations using the indicator</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Data quality</td>
<td>Operation</td>
<td>Data quality refers to representativeness and completeness of the data.</td>
<td>Representativeness is measured by evaluating how well the sample reflects the distribution of cases (for example, in time or place). For instance, data from emergency care departments tends to identify only severe cases, or data based on one emergency department covers only a certain locality. For syndromic surveillance of infectious diseases geographic representativeness is especially important. Completeness is measured by the frequency of unknown or missing data items and also by the frequency of miscoded data.</td>
<td>del Rocio Amezcla et al., 2010 Doroshenko et al., 2005 Josseran et al., 2010 Metzger et al., 2004</td>
</tr>
<tr>
<td>Usefulness</td>
<td>Objective Operation Experience</td>
<td>Usefulness refers to the contribution of the system to the early detection/ reassurance and effective control of an outbreak. An assessment of usefulness addresses the impact or added value by its application.</td>
<td>First, the objectives and priorities of the systems should be reviewed. Usefulness should be assessed by the actions regarding prevention and control taken as a result of the alert. Usefulness can be measured by comparison with the impact of other surveillance systems (for example, did the syndromic surveillance provide sole or additional information that could otherwise not be obtained). Examples can describe how the system has been used for its purpose (and other purposes) including the public health response. Not only early detection but also reassurance can be evaluated (for example, by the number of press conferences or measuring decision maker satisfaction levels).</td>
<td>Buehler et al., 2009 Hulth et al., 2011</td>
</tr>
</tbody>
</table>
### Evaluation Indicators for Evaluation of Syndromic Surveillance Systems (CDC Framework) Continued

<table>
<thead>
<tr>
<th>Evaluation Indicator</th>
<th>Evaluation Focus</th>
<th>Indicator Description</th>
<th>Indicator Measurement</th>
<th>Example of Published Evaluations Using the Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexibility</td>
<td>Objective</td>
<td>The ability of the system to adapt to a change of needs (for example, objective, operating environment, system architecture).</td>
<td>Description of adaptation processes regarding time, personnel and other resources (for example, increase in data providers, modification of case definitions, change in statistical methods, alert thresholds adjustments).</td>
<td>Jefferson et al., 2008 Josseran et al., 2010</td>
</tr>
<tr>
<td></td>
<td>Experience</td>
<td>The ability of the system to adapt to a change of needs (for example, objective, operating environment, system architecture).</td>
<td>Description of adaptation processes regarding time, personnel and other resources (for example, increase in data providers, modification of case definitions, change in statistical methods, alert thresholds adjustments).</td>
<td>Jefferson et al., 2008 Josseran et al., 2010</td>
</tr>
<tr>
<td>Acceptability</td>
<td>Experience</td>
<td>The willingness of people involved in the system to participate (for example, share data).</td>
<td>Acceptability can be measured by:</td>
<td>Jefferson et al., 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acceptability can be measured by:</td>
<td>• the rate of adoptions of the system</td>
<td>Jefferson et al., 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acceptability can be measured by:</td>
<td>• the rate of potential data-providing institutions participating in the system</td>
<td>Jefferson et al., 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acceptability can be measured by:</td>
<td>• the data set completeness</td>
<td>Jefferson et al., 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acceptability can be measured by:</td>
<td>• the timeliness of the steps in the syndromic surveillance process in which humans are involved (for example, manual data collection).</td>
<td>Jefferson et al., 2008</td>
</tr>
<tr>
<td>Portability</td>
<td>Experience</td>
<td>The ability to transfer the system to another setting.</td>
<td>The best measure is the description of examples of transferred systems and the experience with these systems. The system characteristics that might affect portability can also be described.</td>
<td>Jefferson et al., 2008 Josseran et al., 2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The ability to transfer the system to another setting.</td>
<td>The best measure is the description of examples of transferred systems and the experience with these systems. The system characteristics that might affect portability can also be described.</td>
<td>Jefferson et al., 2008 Josseran et al., 2010</td>
</tr>
<tr>
<td>Stability</td>
<td>Experience</td>
<td>Stability refers to the robustness, availability and resilience of the system.</td>
<td>Stability can be measured by the length and consistency of the operation, the frequency of system outages, the length and frequency of downtimes for maintenance, or financial constraints.</td>
<td>Jefferson et al., 2008 Josseran et al., 2010</td>
</tr>
<tr>
<td>Evaluation indicator</td>
<td>Evaluation focus</td>
<td>Indicator description</td>
<td>Indicator measurement</td>
<td>Example of published evaluations using the indicator</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Costs</td>
<td>Experience</td>
<td>Costs can be measured in terms of funding or workload for the different parts, activities or partners of the system. Costs can refer not only to financial and public health costs of false alerts and failures to detect public health events, but also to savings in prevention and control measures.</td>
<td>Costs can be measured as: • direct (for example, data collection, staff costs, software) and indirect (for example, workload for stakeholders not paid for doing syndromic surveillance) • internal and external (for example, data provider) • fixed and variable (for example, response) • differentiated by phase of operation (for example, setup, maintenance) or per year.</td>
<td>Heffernan et al., 2004 Josseran et al., 2010 Kirkwood et al., 2007</td>
</tr>
<tr>
<td>Simplicity</td>
<td>Operation Experience</td>
<td>Simplicity refers to the use of the system and the system architecture.</td>
<td>Measuring simplicity can cover: • the number and characteristic of data sources needed to fulfil the purpose of the system • number of stakeholders receiving reports from the system; level of integration with existing surveillance systems • time spent on different steps (for example data collection, analysis, reporting, qualification and training needs for staff).</td>
<td>Pinheiro et al., 2010</td>
</tr>
</tbody>
</table>
6.4.3 How should the study design be selected?

There are three main study designs used for evaluating syndromic surveillance systems:

- Measuring indicators using qualitative and quantitative methods
- Case studies
- Simulations and exercises.

These are described below. The selection of the study design depends on the focus and objective of the evaluation and there is not one best practice for accomplishing an evaluation.


EXAMPLE 6.1 Evaluation of a national syndromic surveillance system using a mixed methods approach

The syndromic surveillance system in England and Wales, based on data from NHS Direct telephone helplines, was evaluated using a mixed methods approach.

Validity was assessed by comparing the syndromic surveillance data with an established clinically-based surveillance system using a time series analysis (see Figure).

Semi-standardised interviews with eight key stakeholders were undertaken to evaluate usefulness, flexibility, acceptability, portability, stability and system costs. The qualitative assessment was based on the professional judgment of the stakeholders.

NHS Direct and Royal College of General Practitioners Weekly Return
Service time series for Influenza-like Illness by week, England and Wales

Source: Doroshenko et al., 2005.
Case studies As syndromic surveillance systems perform differently for different syndromes, data sources, purposes and contexts it can make sense to focus the evaluation using case studies. Case studies can deal with a particular public health event or one or two selected syndromes (for example, influenza-like illness Doroshenko et al., 2005; Centers for Disease Control and Prevention, 2011) or one data source in a multi-source system (Flamand et al., 2008). Buehler et al. (2009) used real events as starting points for interviews with stakeholders.

**EXAMPLE 6.2 Evaluation of SIDARTHa syndromic surveillance approach for detecting the onset of the Influenza Pandemic 2009**

The SIDARTHa project aimed to develop methods and tools for using syndromic surveillance at regional and local levels across Europe based on routinely collected emergency medical data from three sources. Case studies were used to validate the performance of the system for different purposes (early detection, reassurance of unexpected and expected events) in order to adjust the coding of syndromes, selection and definition of variables of statistical methods, setting of alert thresholds depending on the data source and the syndrome under surveillance, and the user-friendliness of the software tools.

One case study aimed to retrospectively assess the validity and timeliness of detecting the pandemic influenza (A/H1N1) onset in 2009 at the regional level in three European countries (Austria, Belgium and Spain).

The Figures show a comparison between time series and alarms of the syndromic surveillance Cumulative Sum aberration detection algorithm (CUSUM) for influenza-like illness and respiratory illness cases from three emergency care sources with time series and official pandemic periods of the traditional sentinel influenza surveillance in the test regions.

The darker green areas represent the official pandemic periods of traditional influenza surveillance systems. The correlation of time series, sensitivity, specificity, false detection rate and timeliness were evaluated for each data source.

Source: Rosenkötter et al., 2012
Simulations and exercises To test the performance of algorithms and reduce false alarms, simulations are often used. These are combined with the assessment of validity. These methods can be used also to assess the thresholds for signals from syndromic surveillance systems as done by Gault et al. (2009). The more complex application of scenarios to test the performance of algorithms is applied for example by Bork et al. (2006).

EXAMPLE 6.3 Evaluation of ASTER surveillance system applying simulations and exercises

The French syndromic surveillance ASTER allows the monitoring of military populations during a deployment. The system has been used in French Guiana since 2004, in Djibouti since 2006 and by NATO for its international Deployment Health Surveillance Capability since 2010. The system has been evaluated throughout its life, especially using simulations and exercises.

In May 2010, the international capabilities of ASTER were evaluated with a real-size exercise over three days involving the forces of several nations in Kosovo (United States, Germany, Poland, Austria, Czech Republic, France) and the surveillance centre in Munich (Germany). The objective was to assess all human and technical elements of the system, including the surveillance activity. The physicians transmitted their usual outpatient activity, and a nation was required to send additional cases simulating a norovirus outbreak. All activity across the network was monitored, the computer use and the way the epidemiologists monitored the situation was observed and videoed for later analysis. A questionnaire evaluating system usability (Lewis questionnaire, Lewis et al., 1995) and cognitive workload (NASA TLX) was also used.

Detection of the norovirus outbreak by ASTER during the 2010 exercise in Kosovo, in association with an upper respiratory track outbreak in the Czech deployment

There cannot be a gold standard or benchmark for evaluation results. The output and derived recommendations for improvements will be different for different syndromic surveillance activities.

It is important to formulate expected results for the evaluation in the beginning and to interpret the evaluation results in the context of the general and specific objectives of the system under evaluation.

The results of the evaluation should be data that can be used to improve the system and thus the results. The output of the evaluation should feed into concrete and clear recommendations for action, for example, change to more a user-friendly system dashboard to increase portability and simplicity (Pinheiro et al. 2010) or implement regular communication with data providers to increase data quality and timeliness (del Rocio Amezgua et al. 2010).

Recommendations should be an integral part of the evaluation and can function as benchmark for the next evaluation. They should be communicated to all stakeholders.

Comparable evaluation of systems in different countries using a common set of indicators would facilitate understanding and sharing of methods used and problems and successes encountered and revealed by other systems.

In this context, the ECDC is preparing a manual for monitoring data quality and evaluation of surveillance systems that may be relevant for syndromic surveillance purposes.

**Olympic Lessons**

It is important to evaluate all syndromic surveillance systems and services developed for a mass gathering after the event. These evaluations can provide the evidence to support the continuation of legacy systems. It is also important to include syndromic surveillance systems in any overall mass gathering evaluations to demonstrate that syndromic surveillance is an integral part of a national mass gathering response.
### 6.7 Checklist

Consider evaluation at all stages of the life cycle of the system

Define the focus of the evaluation

Select the staff members and stakeholders to be involved in the evaluation

Choose relevant indicators according to the focus of the evaluation

Define appropriate measures for the chosen indicators according to the focus of the evaluation, the system’s general and specific objectives and the expected outcome of the evaluation

Define an appropriate study design according to the indicators and measures, the available resources and the expected outcome of the evaluation

Interpret the results in the context of the general and specific objectives of the system and derive recommendations for action

### 6.8 References


Triple-S project is co-funded by the European Agency for Health and Consumers and the project’s partners. Over the three years from 2010 to 2013 it has brought together 24 organisations from 13 countries (www.syndromicsurveillance.eu).

After creating an inventory of syndromic surveillance activities across Europe, a key goal of the project was to help EU countries create their own syndromic surveillance systems. For this purpose, these guidelines have been developed to aid in the setting up and implementation of syndromic surveillance systems in individual member states at local, regional or national levels.

The guidelines have been written to be as practical as possible. They provide many examples and references of the use of syndromic surveillance in Europe to illustrate each step of the creation and implementation of a syndromic surveillance system.

These guidelines represent the views of many European partners and reflect the experience in Europe. They are not recommending a single European syndromic surveillance system, but a set of common standards that will help exchange and interpret syndromic alerts across Europe, complying with the International Health Regulation and strengthening European health protection.

The Triple-S project’s next step is to suggest a strategy to enable reporting from national syndromic surveillance systems to be comparable at the European level.
### Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aberration</td>
<td>A change in a temporal and/or spatial syndromic indicator that is statistically significant when compared to the expected level.</td>
</tr>
<tr>
<td>Alarm (or signal)</td>
<td>A notification or warning that a statistical aberration represents an epidemiological signal of interest. It represents the operational interpretation of the aberration. It means that the result of the statistical analysis may need further investigation because the population under surveillance is potentially affected by an outbreak or event. The signal may lead to an alarm.</td>
</tr>
</tbody>
</table>
| Alert | A notification or warning issued by the surveillance team stating that the cause of the alarm requires consideration by other parties and/or an intervention. The alert is an important output of the surveillance team highlighting a potentially emerging outbreak or health problem. Depending on the nature of the alarm (for example, severity of the syndrome, magnitude of the alarm) it can result in a range of countermeasures against the health threat. 

**NOTE:** Understanding of the terms aberration, alarm, signal and alert differ between syndromic surveillance systems. Therefore they are difficult to define as independent terms. In general, syndromic surveillance investigations will begin with a statistical aberration before progressing to a signal or alarm and finally an alert. |
| Awareness (situational awareness) | From a public health perspective, situational awareness is one of the main benefits of syndromic surveillance. It helps provide “an accurate, real-time understanding of what is happening on the ground and what options for intervention are feasible”\(^1\). It contributes to knowledge about an epidemiological situation. It relates to the basic epidemiology: the trends, magnitude, geographical areas and people affected by particular syndromes. |
| Baseline | The number of cases observed during the non-outbreak period. It corresponds to what we would expect to see if there was nothing unexpected going on and provides a frame of reference (expected values) that the observed data can be compared to. |
| Companion animals | Domestic-bred animals that are commonly met as companions in the home and are in close relationship with humans, such as dogs or cats. A synonym often used is pets. |

---

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control chart</strong></td>
<td>A control chart is a graphical representation of a statistical process followed over time. It aims to provide an indication of the current evolution of this process, making a distinction between normal and abnormal variation. WA Shewhart(^2) introduced this method for quality control of manufacturing processes in 1931. Today, control charts are a ‘family’ of tools that belong to the statistical field of process control. In syndromic surveillance control charts are used to detect statistically significant variation of a syndrome within a population.</td>
</tr>
<tr>
<td><strong>Countermeasure</strong></td>
<td>Public health and/or medical interventions designed to control or mitigate an outbreak or threat to public health.</td>
</tr>
<tr>
<td><strong>CUSUM chart</strong></td>
<td>The CUSUM chart plots the cumulative sums of the deviations of the sample values (for example syndrome count) from a baseline or benchmark. This sequential analysis technique is an effective method to detect small but persistent departures from the benchmark.</td>
</tr>
<tr>
<td><strong>Disease surveillance</strong></td>
<td>Surveillance, when applied to a disease, means the continued watchfulness over the distribution and trends in the incidence of disease through the systematic collection, consolidation and evaluation of morbidity and mortality reports and other relevant data. (Langmuir AD (1963) The surveillance of communicable diseases of national importance. <em>New England Journal of Medicine</em> 268:182–92)</td>
</tr>
<tr>
<td><strong>Early warning</strong></td>
<td>In the context of syndromic surveillance, this could mean:</td>
</tr>
<tr>
<td></td>
<td>Timely outbreak or event detection while the affected population may be experiencing the early stages of the incident, usually based on pre-diagnostic (or syndromic) data.</td>
</tr>
<tr>
<td></td>
<td>Or</td>
</tr>
<tr>
<td></td>
<td>Detection of an outbreak event in advance of other surveillance systems.</td>
</tr>
<tr>
<td><strong>Epidemiological signal</strong></td>
<td>Time- and/or space-related health statistic or information on the health status of a population. This may be statistical or non-statistical depending on the nature of the surveillance system.</td>
</tr>
<tr>
<td><strong>External factor</strong></td>
<td>External (usually environmental, political, sociological...) factors that may influence the health of the population and consequently influence the fluctuations of the epidemiological indicators.</td>
</tr>
<tr>
<td><strong>GIS</strong></td>
<td>GIS stands for Geographic Information System. It is a system tailored for the representation and analysis of geographical data.</td>
</tr>
<tr>
<td><strong>Health impact assessment</strong></td>
<td>Set of procedures and techniques to judge the potential effects of a policy or intervention on a specific population.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Incidence</strong></th>
<th>The number of new cases of a disease occurring during a given time interval in a specified population or area.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicator</strong></td>
<td>A group of symptoms or signs occurring or a proxy measure in a patient or a population.</td>
</tr>
<tr>
<td><strong>Internal factor</strong></td>
<td>Intrinsic characteristics of the epidemiological indicators, such as secular (long-term) trend, periodicities.</td>
</tr>
<tr>
<td><strong>IT infrastructure</strong></td>
<td>The association of networks, computers, storages, systems and software used in common across all the surveillance system for the acquisition, transmission, storage and processing of data and information.</td>
</tr>
<tr>
<td><strong>Livestock animals</strong></td>
<td>Domesticated animals such as cattle, sheep or goats that are raised on farms for production (commonly meat or milk production).</td>
</tr>
<tr>
<td><strong>Measure of effect</strong></td>
<td>The measure of the impact, usually estimated in terms of excess cases.</td>
</tr>
<tr>
<td><strong>Prospective</strong></td>
<td>Carried out while the event is still in progress.</td>
</tr>
<tr>
<td><strong>Public health investigation</strong></td>
<td>Public health action that aims to confirm, describe and mitigate the effects of an outbreak, event or epidemic. It is organised in a sequence of steps, including preliminary assessment, case definition and identification, descriptive study, hypotheses construction, analytical study, biological verification, control measures and communication.</td>
</tr>
<tr>
<td><strong>Real time (or near real-time) surveillance</strong></td>
<td>Timely disease monitoring that collects and processes surveillance data soon after it has been produced by the source. When the collection and processing of data is done within a few hours (or up to a day) of production, this system is (near) real-time.</td>
</tr>
<tr>
<td><strong>Retrospective</strong></td>
<td>Carried out after the event is over.</td>
</tr>
<tr>
<td><strong>Risk assessment</strong></td>
<td>Characterisation of the threat, estimation of its potential impact on public health, and identification of control measures in order to limit the dissemination and/or impact of the health event.</td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td>Organisation, person or data collection that is able to provide the data or information required for the surveillance system.</td>
</tr>
<tr>
<td><strong>Specific surveillance</strong></td>
<td>Public health surveillance based on the collection and analysis of data related to identified diseases or cases with specific inclusion criteria.</td>
</tr>
<tr>
<td><strong>Standard</strong></td>
<td>A communication protocol or data format with widely disseminated and used specifications.</td>
</tr>
<tr>
<td><strong>Syndrome</strong></td>
<td>A group of symptoms or signs occurring in a patient or a population. A syndrome raises the suspicion of the presence of a disease or condition before it is confirmed.</td>
</tr>
</tbody>
</table>
| **Syndromic surveillance** | Syndromic Surveillance is the real-time (or near real-time) collection, analysis, interpretation and dissemination of health-related data to enable the early identification of the impact (or absence of impact) of potential human or veterinary public health threats that require effective public health action.  
Syndromic surveillance is based not on the laboratory confirmed diagnosis of a disease, but on non-specific health indicators including clinical signs, symptoms as well as proxy measures (for example absenteeism, drug sales, animal production collapse) that constitute a provisional diagnosis (or ‘syndrome’).  
The data are usually collected for purposes other than surveillance and, where possible, are automatically generated so as not to impose an additional burden on the data providers. This surveillance tends to be non-specific yet sensitive and rapid, and can augment and complement the information provided by traditional test based surveillance systems. |
| **Threat** | Any kind of agent, circumstance or event with the potential ability to cause damage to the health of a population. |
| **Visualisation** | A visual, especially graphical, method of data representation that aims to help in the understanding of data. |
| **Z-score** | It is a statistical standardisation process that measures the difference of a value from its mean using the standard deviation as unit of measure. |
[Host organisation name]

and

[data provider]

[title of syndromic surveillance system]:

Information Sharing Agreement
1 Introduction
Syndromic surveillance systems have proved their worth in being able to monitor major public health incidents (for example influenza pandemic) to provide data in real-time relating to health outcomes associated with the incident, or to provide reassurance to healthcare workers involved in treatment.

The [host organisation] is establishing a national surveillance system using [data source] to [aim of system].

1.1 Scope
This Information Sharing Agreement is between the [host organisation] and the [data provider] involved in the [syndromic surveillance system].

This agreement outlines the anonymised dataset to be transferred between the [data provider] and the [host organisation].

This agreement covers all employees, agency workers and volunteers, of all organisations signed up to this agreement.

2 Requirements of the project
The data capture necessary for syndromic surveillance does not impose any extra data capture requirements on [data provider/data source] over and above that which is necessary for good clinical practice.

This agreement will be reviewed and renewed on an annual basis. During this period any proposed changes to the project or dataset will be communicated by the [host organisation] to the [data provider] before any such work is undertaken.

2.1 Anonymised dataset
Encrypted anonymised data will be sent from [data provider] securely to a secure central data warehouse at the [host organisation].

This agreement covers the extraction of the syndromic surveillance dataset. The full list of fields contained within the dataset is presented in the Table.

All data extracted and transferred are anonymised and encrypted using [details of encryption] from point of extraction at [data provider] to point of receipt at [host organisation].

Table: Data fields specified within the syndromic surveillance dataset

<table>
<thead>
<tr>
<th>Data field</th>
</tr>
</thead>
<tbody>
<tr>
<td>List individual data fields</td>
</tr>
</tbody>
</table>
2.2 Operational requirements
This information sharing agreement covers the sharing of an anonymised patient dataset and therefore the majority of confidentiality and data protection issues are not relevant to this agreement.

It is noted, however, that all parties will manage the anonymised data ‘in confidence’ ensuring that existing local mechanisms including data security, confidentiality and ethical assurance is maintained.

3 Agreement

3.1 [Host organisation]
On behalf of the [host organisation] the following authorised signatory agrees to the terms set out in this agreement:

Name:
Designation:
Signature:
Date:

3.2 [Data provider]
On behalf of the data provider the following signatory agrees to the terms set out in this agreement:

Name:
Designation:
Signature:
Date:

4 Annual review
This agreement shall be subject to annual review commencing 12 months from the date of signature on behalf of the Trust in Section 3.2.
Guidelines for designing and implementing a syndromic surveillance system

Deliverable 8, Work Package 6

2013