# Epidemiological data analysis for the early warning alert and response network (EWARN) in humanitarian emergencies

*a quick reference handbook*

World Health Organization
REGIONAL OFFICE FOR THE Eastern Mediterranean

# Epidemiological data analysis for the early warning alert and response network (EWARN) in humanitarian emergencies

## *a quick reference handbook*

World Health Organization
REGIONAL OFFICE FOR THE Eastern Mediterranean

**WHO Library Cataloguing in Publication Data**

Names: World Health Organization. Regional Office for the Eastern Mediterranean

Title: Epidemiological data analysis for the early warning alert and response network (EWARN) in humanitarian emergencies: a quick reference handbook / World Health Organization. Regional Office for the Eastern Mediterranean

Description: Cairo: World Health Organization. Regional Office for the Eastern Mediterranean, 2019

Identifier: ISBN: 978-92-9274-503-5 (pbk.) ISBN: 978-92-9274-504-2 (online)

Subjects: Data Analysis | Disease Outbreaks - statistics & numerical data | Communicable Diseases – epidemiology | Disease Notification – standards | Disaster Planning | World Health Organization

Classification: NLM WA 950

This publication was originally published under ISBN: 978-92-9022-252-1, 978-92-9022-253-8

## Acknowledgements

The WHO Regional Office for the Eastern Mediterranean gratefully acknowledges the technical contribution of the following individuals of the Infectious Hazard Management Programme to the development of this publication: Dr Peter Mala (Medical Officer), Dr Amgad Elkholy (Epidemiologist), Ms Akiko Takeuchi (Technical Officer) and Dr Muhammad Tayyab (Technical Officer).

# Contents

**1. Introduction** ..... **5**

**2. Data collection** ..... **7**
**Role of the surveillance team in data collection and management** ..... **8**
First surveillance level (reporting sources) ..... 8
Intermediate surveillance level ..... 8
Central surveillance level ..... 9
**Types and sources of data** ..... **9**
Alert signal data ..... 9
Weekly aggregate data ..... 11
Outbreak investigation data ..... 11

**3. Analysing EWARN epidemiological data** ..... **13**
**Objectives** ..... **14**
Analysis of the distribution of a disease or public health event ..... 14
Analysis of determinants ..... 14
**Data analysis** ..... **14**
Data quality ..... 15
Performance indicators ..... 15
Recording/entering data ..... 15
Data compilation ..... 16
Data cleaning and preparation for analysis ..... 16
Analysis of weekly aggregate data ..... 16
Analysis of disease outbreak data ..... 20
Summarizing data ..... 26

**4. Data interpretation, feedback and information dissemination** ..... **27**
**Interpretation** ..... **28**
**Feedback and information dissemination** ..... **28**

**Annex: Establishing a disease threshold** ..... **31**

# 1. Introduction

## Background

Humanitarian emergencies often increase the risk of transmission of communicable diseases, resulting in increased morbidity and mortality, particularly from outbreak-prone diseases. To address this increased risk, the World Health Organization (WHO) and its partners established the early warning alert and response network (EWARN), a simplified disease surveillance and response system that focuses on early detection of and rapid response to outbreaks or unusual health events. EWARN is implemented by the ministry of health of the affected country, or its equivalent, with support from WHO and other partner agencies. EWARN is implemented as an adjunct to the national surveillance system during the acute phase of an emergency, when existing communicable disease surveillance systems may be underperforming or disrupted.

Since its introduction in 1999, EWARN has been implemented in response to emergencies in many countries in the WHO Eastern Mediterranean Region, including Sudan (1999, 2004), Pakistan (2005, 2009, 2010), Lebanon (2006), Somalia (2010), the Syrian Arab Republic (2012–2013) and Iraq (2013). It has been adapted to the various settings and has been given different names (e.g. DEWS: Disease Early Warning System; CSR: Communicable Disease Surveillance and Response; EWARS: Early Warning Alert and Response System). However, all these systems are based on the same principles.

## Rationale

This quick reference handbook is intended to guide surveillance staff on their duties and responsibilities in managing and analysing EWARN surveillance data. Skilled human resources are a critical requirement to ensure effective management of EWARN surveillance systems, including generation of reliable and timely data in humanitarian crises. Humanitarian emergency response agencies need timely and good-quality epidemiological information to guide their response activities in order to save lives. However, relevant epidemiological skills, especially good analytical skills for epidemiological data, are rare among health workers in such settings. Although training has been conducted to address these skill gaps, experience has shown that these efforts are often nullified by the high rate of turnover of trained staff in these highly demanding and stressful settings, and short-term nature of the assignments. It is therefore difficult to keep up with frequent training requirements in these settings.

# Aims

This handbook is intended for country-level epidemiologists and surveillance officers who work in data analysis and the production of epidemiological reports for dissemination to stakeholders to inform the need for public health interventions, or the effect of ongoing interventions. Data analysis and dissemination of epidemiological reports is often done at the central level but may also occur at the intermediate and first levels, especially in large hospitals with relevant capacity and the need for this level of epidemiological information. Therefore, the handbook can be used at any level of the local surveillance system depending on local need and capacity. This tool is needed to complement the training these staff receive so as to maintain their skills, especially where staff turnover is high. It is an easy-to-follow, step-by-step reference to show surveillance staff how to manage, analyse, interpret and present epidemiological data in a way that communicates effectively to stakeholders who need the information for public health action. The tool goes a long way to ensuring that trained staff are able to maintain their skills in performing good-quality epidemiological data analysis.

This handbook will equip:

EWARN staff will be equiped with the skills needed to:

- perform EWARN data analysis;
- understand and interpret key epidemiological indicators generated during analysis;
- use EWARN data to describe the health situation and detect outbreaks early; and
- identify risk factors and causes to guide optimal interventions.

The handbook will also equip surveillance officers with the necessary skills to present epidemiological information and reports in an understandable and meaningful way to stakeholders, including public health actors and policy-makers who may not have an epidemiology background.

The handbook is organized in three parts. The first part covers data collection: the role of the surveillance team in data collection and management of sources of surveillance data, and how to collect data and ensure their quality. The second part deals with analysing EWARN epidemiological data: the methods, tools and steps of data analysis and presentation. Important epidemiological health indicators that should be generated on a regular basis are also discussed. The third part discusses data interpretation and information dissemination: the different ways of sharing epidemiological information with all surveillance levels, stakeholders, and all those who need to know. An annex is also included which outlines how to set disease thresholds.

# 2. Data collection

## Role of the surveillance team in data collection and management

There are specific skills, resources and roles for EWARN officers at each of the three levels (first, intermediate, and central).

- The first surveillance level focuses on proper data recording and timely sharing.
- The intermediate level focuses on cleaning and aggregating data, and assisting the first level with issues.
- The central level focuses on conducting in-depth epidemiological analysis, and supporting the first and intermediate levels with any issues.

## Types and sources of data

There are three main types of data in an EWARN, each with its own purpose, mode of collection, and action points.

- Alert signal data: unstructured informal information that signals a health event or potential risk.
- Weekly aggregate data: structured and systematic collection of data, used to calculate health indicators.
- Outbreak investigation data: which ultimately serves to control outbreaks and reduce subsequent morbidity and mortality.

# Role of the surveillance team in data collection and management

Responsible focal points at the first level (i.e. reporting level) should be trained in surveillance procedures, including what should be reported, when it should be reported, to whom and how (i.e. using which form or data transmission methods). The focal points at all levels should be familiar with the diseases under surveillance, their case definitions and alert and epidemic thresholds, and the role of the laboratory in surveillance. They should also be able to collect samples, and package and coordinate their transportation to the laboratory. In order to detect any unusual changes in disease occurrence patterns early, so as to ensure early identification of outbreaks, focal points at intermediate and higher levels are also expected to review and analyse surveillance data on a regular basis: by time in order to detect any unusual trend patterns; by place to identify any clustering; and by person for any possible association with demographic risk factors.

## First surveillance level (reporting sources)

**Skills and essential resources**

- Should be well informed about surveillance reporting procedures, diseases under surveillance and reporting forms
- Should have a sufficient supply of data recording (i.e. registers) and reporting (i.e. reporting forms) tools.

**Roles and activities**

- Ensure that a proper record or register of health facility consultations is maintained at the health facility.
- Enter data in an electronic form (if applicable) or record data in appropriate reporting forms, and regularly transfer data electronically or in paper form to the second surveillance level on a timely basis.
- Document and immediately report any unusual events detected by observation or through informal sources to the intermediate surveillance level through the alert system.
- Maintain a detailed line list of suspected cases of the disease being monitored during an outbreak.
- Perform basic analysis and calculate health indicators for specific diseases (if required by the health facility administration).

## Intermediate surveillance level

**Skills and essential resources**

- Should have a certificate/diploma or higher of epidemiology training
- Should be equipped with IT equipment for data management.

**Roles and activities**

- Receive and document receipt of the completed aggregate data/case reporting forms or electronic data from the reporting sources.
- Verify alert signals and reports, and maintain the records in an alert logbook.
- Review a random sample (~10%) of the data for accuracy and completeness (data validation), and provide feedback to the reporting sources.
- Clean up data before weekly transfer of data to the central level.
- Contact less active and silent reporting sources for follow-up and problem-solving.
- Compile all data received from reporting sources in paper form or an electronic database.
- Transfer data weekly to the upper level/central level in paper form or electronically and document the transfer.
- Undertake basic data analysis for all reportable diseases by frequency and rate.
- Identify any unusual event or increase in the number of cases of reportable diseases by comparing the numbers to the specific disease threshold.

- Verify and validate any suspected unusual event received through informal sources, and report it to the central level for evaluation.
- Provide feedback to lower level/reporting sources.

## Central surveillance level

**Skills and essential resources**

- Should have at least a master's degree training in public health.
- Should be equipped with IT equipment for data management.

**Roles and activities**

- Maintain and analyse data from the different surveillance levels.
- Receive, validate and maintain surveillance data from lower levels on a weekly basis.
- Compile data from all the country into one database every week.
- Contact surveillance teams at lower levels to ensure data is received regularly.
- Evaluate alerts received to detect any aberrations.
- Ensure quality of data and contact subnational levels to validate data if required.
- Conduct in-depth epidemiological analysis (disease trends, high-risk groups and high-risk areas) from the national surveillance data.
- Prepare epidemiological reports for dissemination to stakeholders.
- Work closely with laboratories to ensure samples are collected, packaged and transported to the laboratory promptly and results are obtained.

# Types and sources of data

Before any analysis can take place, epidemiological data are needed. Data for epidemiological analysis are obtained from the EWARN surveillance system based on a scheduled reporting frequency. While the focus of this handbook is on EWARN surveillance data, the same analytical principles, methods and processes apply to other sources of surveillance data.

EWARN surveillance systems provide data on an ongoing basis from designated health facilities, and field medical units among refugees and displaced populations at a predetermined reporting frequency, usually weekly for priority epidemic-prone diseases. The system also includes immediate notifications of suspected outbreaks and unusual health events from the reporting sites and sources outside the health system, such as the community and the media.

The number of priority diseases and parameters (or variables) required for each disease should be kept to an essential minimum to ensure simplicity of the EWARN surveillance system. The EWARN system is not designed for complete accounting of disease incidence in the community at large, but rather to enable rapid detection of epidemics among those affected by humanitarian emergencies. A core set of variables should include case counts, two age categories (0–4 years and 5 years and older) and applicable laboratory information, if available. Additional information may be useful for programme planning (e.g. sex information or narrower age categories) but this will not influence public health response efforts in the event of an outbreak during an acute emergency.

## Alert signal data

An alert is an unusual health event that can signal the early stages of an outbreak or a public health event, such as a chemical poisoning incident. It consists mainly of unstructured informal information about health events or risks that may represent an acute risk to human health. The main aim of collecting alert signals is to detect a single or a cluster of cases or deaths in the same time and area. The information collected is varied and originates from multiple, often not predetermined sources, both official and unofficial, including rumours and informal reports. Data should be reported immediately as soon as received by official health workers. The following actions should be taken.

- At the health facility/reporting source, after receiving data from informal sources, immediately notify the second level.
- Collect as many details as possible about the alert: when, where, who, response initiated, outcome and mapping, as appropriate.
- Ask about similar cases in the contacts, neighbourhood and nearby health facilities.
- Write your preliminary interpretation of the data collected, e.g. case–fatality rate is xx%, cases are mostly from xx village, unusual number of cases for this time of the year, the increase in cases of the disease could be related to the animal breeding season.
- Share the data collected as soon as possible with the next surveillance level. Do not wait until you have collected all the data; share what you have collected daily with the second level.

**Alert reporting sources**

*Health sector source*

- reporting of one, or more, unusual health event by a community health worker
- observation of a single event or cluster of events by clinicians in a health facility, or by laboratory staff or surveillance officer
- review of routinely collected surveillance data for unusual health event(s) or trend patterns.

*Non-health sector sources*

- the media
- informal networks
- community
- internet websites
- private sector
- nongovernmental organizations.

**Verification of an alert**

Many alerts are detected and the majority may not end up being outbreaks. Thus, every alert must be verified to determine whether or not it is a false alert. False alerts could be due to:

- error in reporting, or laboratory or diagnostic error
- increased concerns about a disease
- new health provider
- improvement in diagnostic methods
- changes in local reporting practice
- changes in case definition
- population increase (migration)
- increased access to health care
- batch reporting.

**False alert exclusion**

False alerts should be excluded by collecting information from all available sources. Information to collect to verify an alert might include:

- case definitions used and method of diagnosis and laboratory testing
- symptoms and signs of cases to verify the diagnosis
- case-management details of cases and deaths, and the number of health care staff affected
- date of onset of the first and the most recently detected cases
- place and date of consultation or hospitalization
- age, sex and vaccination status of patients, where relevant
- place of residence at onset of illness
- clustering: geographical, personal and time relationships between cases.

### Managing alert signal data

Any alert received should be recorded in an alert monitoring logbook where time, action taken and result of investigating the alert (possible outbreak or false alert) are recorded. No data analysis is required for this type of data received.

## Weekly aggregate data

This is a structured and systematic collection of data used to calculate health indicators. These data are usually produced from a number of well-identified, mostly health-based formal sources. Data are collected according to established case definitions which are either disease-specific (e.g. malaria) or syndromic (e.g. acute watery diarrhoea). Immediate case reporting is required for some diseases in accordance with the country policy, or weekly reporting of aggregate data is required for other diseases based on the urgency of interventions and investigations.

### Preparing weekly aggregate data

i. Describe individual cases: the first step in preparing the aggregate weekly data is to collect basic information about each case. Minimal data are collected for the sake of simplicity and usefulness of data.

ii. Aggregate the data: individual case data are summarized and analysed to describe the situation and to determine unusual disease patterns in comparison with baseline values and thresholds.

### Sources of weekly aggregate data

- health care facilities
- health professionals
- laboratories
- animal health data
- entomological data.

### Managing weekly aggregate data

Aggregate data can be recorded in paper form or entered in an electronic form such as Microsoft Excel or in statistical software. The data are used to detect any change from the usual disease pattern in order to detect outbreaks early. Details of the methods of setting disease thresholds for aberration detection are given in the annex.

## Outbreak investigation data

The main objective of an outbreak investigation is to control the outbreak and thus reduce morbidity and mortality. The investigation should begin as soon as an alert is detected and an outbreak verified. Once an outbreak is thought to be occurring, surveillance data for the specific disease concerned are collected on a daily basis using a line list. This is more detailed and is essential for monitoring the evolution of the outbreak; the line list is also useful for identifying possible causes of an outbreak and who is at risk so that control measures can be implemented. It differs from alert verification (determining whether an outbreak is actually occurring) but is a continuation of the process of characterizing an unusual health-related event in which there is an increase in cases above an expected baseline.

### Sources of outbreak data

- epidemiological data on the cases affected by the disease causing the outbreak
- epidemiological data on controls not affected by the outbreak disease (for the case–control study)
- environmental data (in some diseases, e.g. typhoid and cholera)
- data on the vector (in some diseases, e.g. Rift Valley and yellow fever)
- laboratory data.

**How to collect outbreak data**

During an outbreak, you need to collect epidemiological data for all cases suspected of having the disease to better describe the outbreak. A line list is the appropriate tool to collect this type of data on one sheet or one database so you can analyse and interpret the data easily. Laboratory data should be included in the line list for confirming the diseases.

**Managing outbreak data**

Epidemiological outbreak investigation data should be entered/recorded and analysed on a daily basis to obtain more information on the outbreak situation. Data should be analysed by time, place and person, in addition to the cause of outbreak and its risk factors, if applicable. The line list can be recorded as a table in paper form or using electronic database software (e.g. Excel). Other types of outbreak data, e.g. environmental and vector data, are used to guide intervention and control measures.

# 3. Analysing EWARN epidemiological data

## Analysing EWARN epidemiological data

EWARN data can be analysed by person, time and place:

- Person, based on demographic characteristics of the population.
- Time, based on temporal units such as weeks, but also seasons.
- Place, or geographical distribution.

## Data analysis

Analysis based on quality EWARN data is crucial to inform public health interventions. For proper data analysis:

- monitor data quality at all stages, by following a predefined set of indicators;
- follow proper data entry protocol and best practices;
- compile and clean data from different sources;
- conduct proper analysis of weekly aggregate data and disease outbreak data, including tabulation and indicator calculation;
- highlight and summarize key points like aberrations from the normal pattern, start and end date of possible outbreaks, clustering, and other trends and patterns in the data.

# Objectives

The aim of epidemiological data analysis is to determine the distribution and determinants of health-related events (including disease), and use the findings to control diseases and other health problems.

## Analysis of the distribution of a disease or public health event

Analysis of distribution is linked to the descriptive aspect of the epidemiological analysis. It includes presentation of the data by person, time and place.

**Person**

Analysis is done by demographic characteristics of the population covered by the EWARN surveillance system. Commonly used characteristics include age and sex. Other characteristics that may be considered, depending on the disease of interest, include socioeconomic characteristics such as education, marital status, income level, type of employment and behavioural characteristics, e.g. smoking. Description of the occurrence of the disease of interest according to demographic characteristics can be presented in tables, graphs/figures and narrative text.

**Time**

Time can be presented as day of the week, week of the month, month of the year or year. Time also implies seasons of the year (cold, hot, rainy, dry, winter or summer), depending on how rapidly new cases occur or the rate of transmission. Description of the occurrence of the disease by time, or trend analysis, can be presented in trend graphs/figures, tables and narrative text.

**Place**

Description of data by place, also referred to as geographical distribution of cases, enables the determination of pace, direction and spread of the disease; how cases are distributed; and whether there is any clustering in a particular place. Spatial distribution analysis of cases can be presented using maps, tables and narrative text.

## Analysis of determinants

The study of determinants involves inferences about potential causal relationships concerning the observed distribution patterns, and the application of statistical models to determine the significance of the associations.

# Data analysis

Epidemiological data in a raw state are just long rows and columns of numbers and text that do not convey much in terms of key questions about which disease, or event, is of public health concern: where is it occurring, when did it start, who is at risk, why is it occurring, how did it start, what is the current status, how serious is it, what is being done or should be done about it. The purpose of epidemiological analysis is to translate these numbers and texts into information that is understandable and informative to key stakeholders, including policy-makers and the affected community, in order to ensure that well-informed and effective curative, preventive and overall control measures are taken against the disease, or public health event, of concern.

The aim of analysing and interpreting EWARN data is to inform public health interventions in the appropriate time. Immediate notification from reporting facilities upon detection of alerts (originating from health-facility or community rumours) is the primary method of detection of most outbreaks. In addition, whatever the alert mode, analysis of and feedback on the outcome of the alert verification process and weekly aggregate data are critical for continuity of information flow and to maintain the reporting process. After preparing your data, now you are ready to do your data analysis. How you do this will depend on the type of data available as either aggregate data or a line list.

The EWARN data to be analysed will be received in the form of immediate alerts of unusual health events or suspected outbreaks, or alert logs/weekly aggregated reports, or line lists in the case of confirmed outbreaks. Irrespective of the type of data, the process of epidemiological data analysis involves a series of logical steps. Before starting your data analysis you need to prepare the data for the analysis as described below.

## Data quality

It is important to monitor data quality at all stages of managing data, i.e. during data collection, data transfer, data cleaning and analysis. Good-quality data ensure that the collected data can meet the objectives of the EWARN system. The quality of the data collected at reporting sources will determine the quality and usefulness of the data at all stages of the reporting process. Monitoring quality also helps to improve the analysis and interpretation of EWARN data at all levels. It is also very important for data to be available to those who need it at the appropriate time for early outbreak detection, implementation of effective interventions and disease control.

Specific indicators are used to monitor data quality at all levels. Examples of indicators that could be used to evaluate data quality are given in the following section.

## Performance indicators

Indicators for monitoring the performance of an EWARN surveillance system and how to calculate them are described below.

- Completeness of data recording: number of filled surveillance forms that have all variables completed/all filled forms × 100
- Completeness of case reporting: number of cases in filled surveillance forms for each reportable disease/ number of cases of the disease registered in the outpatient records × 100
- Completeness of alerts recording: number of alerts recorded in the logbook/number of reported alerts in health facility records × 100
- Timeliness of alert reporting: number of alerts reported within 24 hours of detection/total number of alerts × 100
- Completeness of weekly reporting: number of reporting sources that reported data last week/all reporting sources × 100
- Timeliness of weekly reporting: number of reporting sources that reported on time last week/all reporting sources × 100
- Timeliness of alert verification: number of alerts verified within 24 hours/total number of alerts × 100
- Timeliness of outbreak investigation: number of outbreaks investigated within 72 hours/total number of outbreaks × 100

## Recording/entering data

The first step in data preparation is to have your data in a form that can be analysed. Data should be extracted from the health facility registers into reporting forms, and from the reporting forms data should be tabulated into weekly aggregate data for weekly reports, or line list tabulation for outbreak data. This should be done at the health facility or first surveillance level. This can be done using management or statistical software (e.g. Excel, Epi Info, STATA).

If an electronic database is being used, keep your database up to date and clean, and have a back-up copy and update it regularly. Use different media to maintain your data so you do not lose them. Always keep a list (data dictionary) with the variables, their abbreviations and the meaning of the codes under each variable in case anyone working with data needs to use it. The quality of data should be taken care of during data entry or recording.

## Data compilation

In a typical early warning surveillance system, district levels regularly receive data from reporting sites, as well as additional data on unusual health events from non-health care settings. Once received, the data from the different sources need to be compiled in one database to allow data analysis. Compilation can be done by merging data files from different sources if submitted electronically, or done manually by entering the data from the different reporting sources into one table or line list.

## Data cleaning and preparation for analysis

Before starting data analysis, every effort should be made to ensure the quality of the data as follows.

- Ensure at least 95% of the cells are complete with the correct information.
- Ensure there are no discrepancies in the data, e.g. a child who is married or a male who is pregnant. If you find discrepancies, revise the original forms, talk to data collectors or search for other data sources. Then make the corrections required and keep the original file.
- Ensure that age is in an appropriate format and range, i.e. no subject age > 100 years and age in months only for infants < 1 year.
- Ensure laboratory results are logical, e.g. a patient with meningococcal meningitis confirmed by culture whose cerebrospinal fluid is clear is not logical.
- Develop new variables if needed in new columns. For example, length of hospital stay (date of discharge – date of hospital admission), or categorize patient age into five or six categories according to the disease type.

## Analysis of weekly aggregate data

EWARN surveillance systems produce three types of information – alert signals data, outbreak data and weekly aggregate data.

Alerts require simple descriptive analysis in terms of number and type of alerts received, time to verification (i.e. proportion of alerts verified within 24 hours), number of verifications done, and status of the alerts (still open and further action is pending; possible outbreak for further investigation; false alarm; and considered closed).

Outbreaks of the weekly reported diseases can be detected from analysis of weekly aggregate data for aberration detection. This should be done at the subnational level which has aggregate data from different sources and is capable of doing basic data analysis and comparisons by time and place.

Before aggregate data analysis, data from different sources should be sent to the subnational level on a weekly basis, either electronically or in paper form, as case-based data or in an aggregate form or both. At the subnational level, data should be compiled in one database either electronically, e.g. using Excel or Epi Info software, or manually in table form. After data compilation, the data should be checked for completeness, discrepancies and soundness. Reporting sources should be contacted to report any substandard data quality and to correct any incomplete or illogical data.

Example A gives an analysis of aggregated data.

### Data tabulation

Tabulate the number of cases for the epidemiological week 4 of 2016 (Table 1). Tabulation of disease occurrence patterns by age and place provides an overview of the data and makes it easier to spot unusual numbers in terms of who is affected (e.g. age group), and where they are coming from (e.g. place/districts).

### Calculation of epidemiological indicators

The aim is to describe the current situation and detect early any aberration in the disease occurrence in specific areas and weeks. Calculate the incidence for each reportable disease and compare between different districts.

**Table 1. Disease occurrence pattern by district and age in week 4, 2016**

| Disease | Age group (years) | District A | District B | Total |
|---|---|---|---|---|
| Acute diarrhoea | < 5 | 12 | 19 | 31 |
| | ≥ 5 | 5 | 28 | 33 |
| Leishmaniasis | < 5 | 0 | 0 | 0 |
| | ≥ 5 | 1 | 0 | 1 |
| Malaria | < 5 | 2 | 0 | 2 |
| | ≥ 5 | 4 | 0 | 4 |
| Measles | < 5 | 0 | 15 | 15 |
| | ≥ 5 | 0 | 1 | 1 |
| Meningitis | < 5 | 0 | 1 | 1 |
| | ≥ 5 | 1 | 7 | 9 |
| Pertussis | < 5 | 1 | 2 | 3 |
| | ≥ 5 | 0 | 2 | 2 |

### Example A: analysis of weekly aggregate data

You are the surveillance officer of country X. The country has a small population of 100 000 people. It has two districts, District A with a population of 45 000 and District B with a population of 55 000. The national surveillance system has eight primary care health facilities in District A, 11 primary care facilities in District B and a secondary district hospital in each of the two districts.

The EWARN system has standard tools for reporting, including immediate notification forms, weekly reporting forms and national public health surveillance guidelines. The system has only six priority communicable diseases based on local disease epidemiology: diarrhoeal diseases, leishmaniasis, malaria, measles, meningitis and pertussis.

Health facilities are responsible for immediate notifications through designated telephone and fax numbers and e-mail, and also for weekly reporting. Weekly reporting is based on standard epidemiological weeks. All the immediate notifications are received by the district surveillance office and compiled in alert logs for monitoring and further action. Weekly reports are received by the district surveillance office for aggregation in preparation for district level analysis and for onward transmission of district aggregate data to the national (central) level.

Weekly aggregated surveillance data for the first four weeks of 2016 are provided to illustrate data analysis at the central level (See Table 2). Similar analysis can be replicated at intermediate levels.

Comparing the incidence rate for several weeks is very informative and shows the trend and identifies any unusual increase. These indicators include:

*Incidence*: number of new cases of a specific disease during a specific week/number of population during the specific week (if population at risk is known). However if the population at risk is not available, proportional morbidity (see item below) is often used as an alternative measure of incidence.

The population of the country is 100 000, categorized as: 40 000 males and 60 000 females; 30 000 aged < 5 years and 70 000 aged ≥5 years; and 45 000 living in District A and 55 000 in District B.

Therefore, the national incidence of measles in country X in week 4, 2016:
= (16/100 000) × 100 000 = 16/100 000 population.

*Proportional morbidity*: number of cases of a certain disease/total number of consultations in a certain week and place × 100.
This is used to describe trends and identify possible aberrations from the norm when the population denominators are unknown or changing. A graph showing the proportional morbidity rate for several weeks is informative to compare the rates and show the trend.

Case–fatality rate: number of deaths caused by a specific disease/total number of cases of that disease × 100
In this illustration above, there were five deaths due to measles in week 4 (Table 2):
= 5/16*100
= 31.25%

**Table 2. National weekly aggregated data, weeks 1–4, 2016**

| District | Year | Week | Disease | < 5 years | ≥ 5 years | Males | Females | Deaths | Total cases |
|---|---|---|---|---|---|---|---|---|---|
| DISTRICT-A | 2016 | 1 | Leishmaniasis | 0 | 2 | 0 | 2 | 0 | 2 |
| DISTRICT-A | 2016 | 1 | Measles | 4 | 0 | 2 | 2 | 2 | 4 |
| DISTRICT-A | 2016 | 1 | Pertussis | 5 | 2 | 3 | 4 | 0 | 7 |
| DISTRICT-B | 2016 | 1 | Acute diarrhoea | 3 | 5 | 4 | 4 | 2 | 8 |
| DISTRICT-B | 2016 | 1 | Leishmaniasis | 0 | 0 | 0 | 0 | 0 | 0 |
| DISTRICT-A | 2016 | 2 | Acute diarrhoea | 12 | 23 | 15 | 20 | 2 | 35 |
| DISTRICT-A | 2016 | 2 | Meningitis | 1 | 1 | 1 | 1 | 1 | 2 |
| DISTRICT-A | 2016 | 2 | Meningitis | 2 | 2 | 1 | 3 | 0 | 4 |
| DISTRICT-B | 2016 | 2 | Measles | 0 | 2 | 1 | 1 | 1 | 2 |
| DISTRICT-B | 2016 | 2 | Pertussis | 0 | 2 | 1 | 1 | 0 | 2 |
| DISTRICT-A | 2016 | 3 | Acute diarrhoea | 12 | 12 | 10 | 14 | 2 | 24 |
| DISTRICT-A | 2016 | 3 | Pertussis | 0 | 0 | 0 | 0 | 0 | 0 |
| DISTRICT-A | 2016 | 3 | Leishmaniasis | 0 | 1 | 1 | 0 | 0 | 1 |
| DISTRICT-B | 2016 | 3 | Malaria | 0 | 0 | 0 | 0 | 0 | 0 |
| DISTRICT-B | 2016 | 3 | Measles | 2 | 0 | 1 | 1 | 1 | 2 |
| DISTRICT-A | 2016 | 4 | Acute diarrhoea | 12 | 5 | 10 | 7 | 0 | 17 |
| DISTRICT-A | 2016 | 4 | Malaria | 2 | 4 | 4 | 2 | 3 | 6 |
| DISTRICT-A | 2016 | 4 | Meningitis | 0 | 1 | 0 | 1 | 0 | 1 |
| DISTRICT-A | 2016 | 4 | Measles | 0 | 0 | 0 | 0 | 0 | 0 |
| DISTRICT-A | 2016 | 4 | Pertussis | 1 | 0 | 0 | 1 | 0 | 1 |
| DISTRICT-A | 2016 | 4 | Leishmaniasis | 0 | 1 | 1 | 0 | 0 | 1 |
| DISTRICT-B | 2016 | 4 | Malaria | 0 | 0 | 0 | 0 | 0 | 0 |
| DISTRICT-B | 2016 | 4 | Leishmaniasis | 0 | 0 | 0 | 0 | 0 | 0 |
| DISTRICT-B | 2016 | 4 | Acute diarrhoea | 19 | 28 | 20 | 27 | 3 | 47 |
| DISTRICT-B | 2016 | 4 | Meningitis | 1 | 7 | 5 | 3 | 3 | 8 |
| DISTRICT-B | 2016 | 4 | Measles | 15 | 1 | 12 | 4 | 5 | 16 |
| DISTRICT-B | 2016 | 4 | Pertussis | 2 | 2 | 1 | 3 | 0 | 4 |

### Analyse the data by time

This is usually done to characterize the trend over time and detect any aberration from the normal situation.

The first analysis is usually a comparison of the number of cases reported for the current week with the numbers in the preceding weeks. An abrupt increase or a gradual build-up in the number of cases can be detected by looking at the table or graph. It is particularly informative when new cases are reported promptly and regularly.

In this example, data are available for the first four weeks of 2016. In reality, you need to present data starting from week 1 of any particular year up to the last reporting week. In order to detect aberrations, you also need to compare the current year's weeks with the same weeks in the previous year, if you have these data available, as illustrated in Fig. 1.

Fig. 1 indicates that something unusual occurred from week 2 in 2016 as there was a substantial increase in new cases of acute diarrhoea and the incidence remained more than twice as high as in the two previous years.

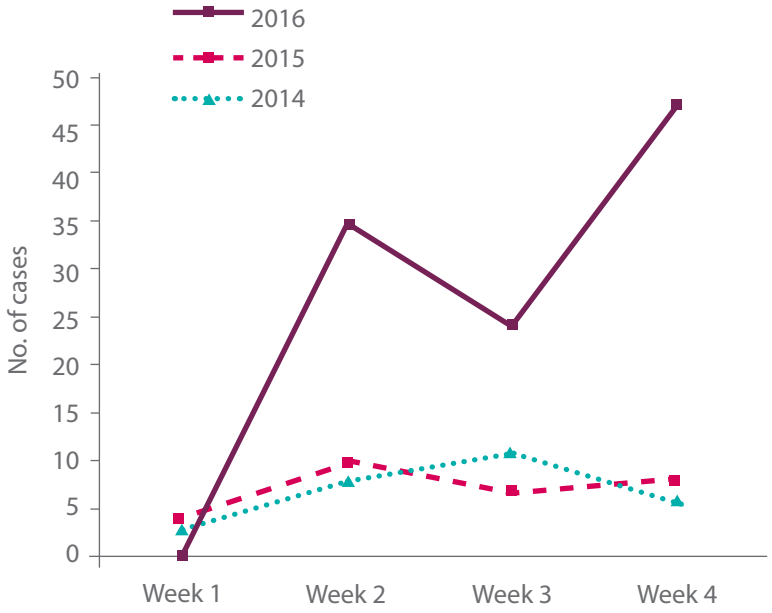



**Fig. 1. Trend in acute diarrhoea in weeks 1–4 in 2016 compared with the same period in the previous two years**

You can also show the trend of multiple diseases of interest (maximum four diseases) in the same graph as shown in Fig. 2.

### Analyse data by place

Ideally, we would like to know where people were exposed but, more commonly, all we know is where the people affected live or work. Place data are usually shown with a map or table containing the number of cases in each area (e.g. village, district, governorate or locality). The incidence rate is a more accurate measure if population data are available.

You can also display the incidence of a disease on a map using different colours. Usually a dark colour indicates a high incidence. You can also write the incidence in numbers on each district to show the areas with higher incidence rates of a certain disease (Fig. 3).

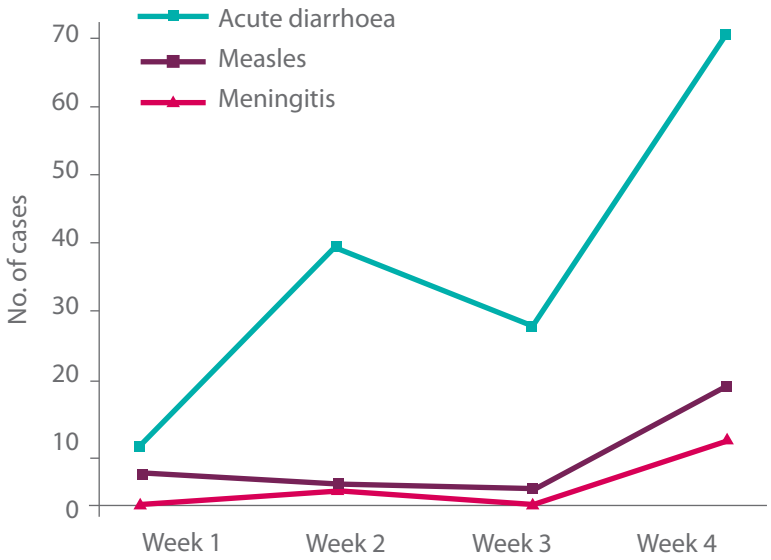



**Fig. 2. Trends in the leading causes of morbidity in weeks 1–4, 2016**

### Analyse data by person

When analysing person data, we need to know the people who are most affected, e.g. age group, sex.

Sometimes, we can have information on what cases may have in common, but this may not be known until further investigations have been conducted. Calculate rates using appropriate population denominators, if possible, to identify the groups with a higher risk of getting the disease (risk group). Develop a table and graph with the number of cases for each disease in each age group as illustrated for acute diarrhoea cases in weeks 1–4 (Fig. 4).

## Analysis of disease outbreak data

A disease outbreak is an increase in the number of cases above the epidemic threshold (i.e. occurrence of a disease at levels exceeding known endemic levels of the disease at the place and time concerned). When an outbreak is confirmed, then a detailed line list of cases is usually prepared at the health facility level and submitted daily to responsible epidemiologists at intermediate and/or central levels. The line lists are consolidated at higher levels as shown in Example B.

### Analyse the data by place

Prepare a map with the relevant landmarks for the disease and then add dots that represent the number of cases in each area. A spot map is useful in outbreak investigations as it can show clustering of cases in certain areas or a relation between a certain risk factor and the presence of cases. This can be done manually using a blank map, as in the following example, or using mapping software if you can identify the Global Positioning System (GPS) coordinates of each case.

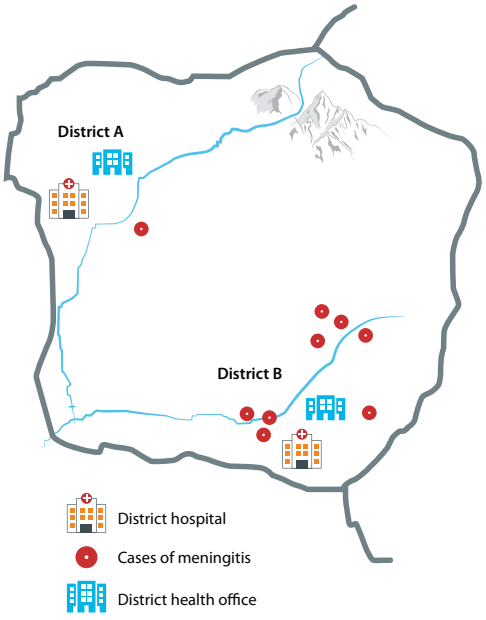



**Fig. 3. Dot map of confirmed meningitis cases in Districts A and B in Country X**

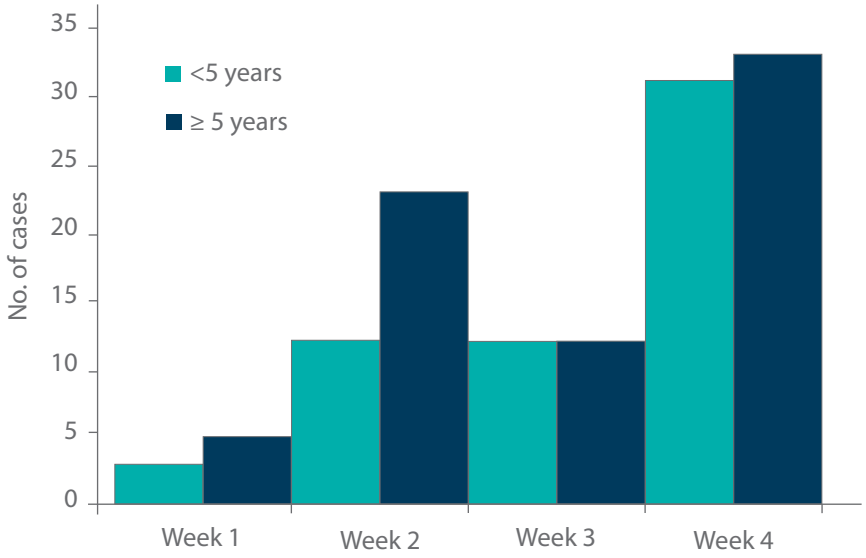



**Fig. 4. Trend in acute diarrhoea cases by age for weeks 1–4, 2016**

### Example B: analysis of outbreak data

Country X is divided into two districts (A and B). Since the first suspected case of hepatitis A was reported on 1 January 2016 in the western part of the country, the number of cases has reached 25, including five deaths, as of 12 January 2016. The main clinical features of the disease are fever and jaundice, the majority of cases appear to be those who rely on the river as their main source of water. All suspected cases were referred to the district hospital of district A.

The nine affected villages have a total population of 600 people. The infectious diseases hospital has received all the cases and maintained a comprehensive line list of all of them (Table 3). You are now responsible for designing the optimal outbreak management plan. You are required to conduct a formal study of the outbreak by analysing the case data that the hospital has made available to you.

Key questions for any outbreak investigation that you need to bear in mind include:

- What is the disease?
- Who is at risk of becoming ill?
- Where is the source of the epidemic?
- What type of epidemic is this?
- What is the route of transmission?
- What measures should be implemented to control the outbreak?

The map (Fig. 5) shows that the hepatitis A cases are concentrated in the western villages of District B and two neighbouring villages in District A. There is significant clustering in two adjacent villages: v1 (nine cases) and v10 (six cases) that sit along the river.

#### Analyse outbreak data by person

To identify the risk groups, you need to analyse the data by all personal characteristics you have collected in your line list using tables or graphs (Table 4).

**Table 3. Line list of confirmed hepatitis A cases in Districts A and B in Country X**

| Patient | Residence (village) | Sex | Age (years) | Date of onset (m/d/yyyy) | Outcome |
|---|---|---|---|---|---|
| 1 | v3 | F | 45 | 1/1/2016 | Recovered |
| 2 | v2 | M | 46 | 1/2/2016 | Admitted |
| 3 | v10 | F | 25 | 1/3/2016 | Died |
| 4 | v1 | F | 50 | 1/4/2016 | Admitted |
| 5 | v2 | M | 28 | 1/4/2016 | Died |
| 6 | v10 | M | 27 | 1/4/2016 | Died |
| 7 | v5 | M | 29 | 1/4/2016 | Recovered |
| 8 | v4 | M | 30 | 1/5/2016 | Admitted |
| 9 | v10 | F | 28 | 1/5/2016 | Recovered |
| 10 | v1 | F | 40 | 1/5/2016 | Recovered |
| 11 | v15 | F | 28 | 1/5/2016 | Recovered |
| 12 | v1 | F | 26 | 1/7/2016 | Recovered |
| 13 | v6 | F | 27 | 1/7/2016 | Recovered |
| 14 | v10 | F | 39 | 1/7/2016 | Recovered |
| 15 | v1 | M | 19 | 1/7/2016 | Recovered |
| 16 | v8 | M | 23 | 1/7/2016 | Recovered |
| 17 | v10 | M | 22 | 1/8/2016 | Recovered |
| 18 | v1 | M | 30 | 1/8/2016 | Admitted |
| 19 | v9 | M | 32 | 1/9/2016 | Recovered |
| 20 | v1 | F | 19 | 1/11/2016 | Recovered |
| 21 | v10 | M | 25 | 1/11/2016 | Recovered |
| 22 | v1 | M | 33 | 1/12/2016 | Died |
| 23 | v1 | F | 47 | 1/12/2016 | Recovered |
| 24 | v9 | M | 50 | 1/12/2016 | Died |
| 25 | v1 | M | 22 | 1/12/2016 | Admitted |

**Analyse by time (i.e. prepare an epidemic curve)**

In an outbreak, an epidemic curve is the most appropriate form to describe the number of cases by time. It is a type of histogram used to depict the time course of an outbreak. It provides a simple visual display of the magnitude and time trend of an outbreak.

The epidemic curve shows:

- the magnitude of the epidemic over time to distinguish epidemic from endemic disease;
- where you are in the course of the epidemic, i.e. still on the rise, on the decline, or after the epidemic has ended to predict the number of cases you will see;

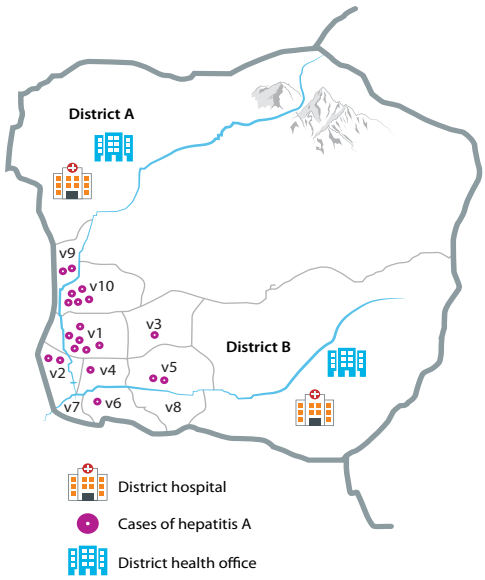



**Fig. 5 Dot map of confirmed hepatitis A cases in Districts A and B in Country X**

- how long it took to identify a problem and the effect of the intervention measures;
- the way the disease is transmitted in the population, e.g. point source, intermittent or propagated outbreak.

To draw an epidemic curve for hepatitis A in Country X, first summarize the data from the line list (Table 3) by date of onset of hepatitis A cases as shown in Table 5.

Then, if automated data analysis applications are not being used and Excel is being used, highlight the data and select insert chart and label the chart accordingly (Fig. 6).

**Table 4. Characteristics of confirmed hepatitis A cases and deaths**

| Characteristic | Cases (*n* = 25) | Deaths (*n* = 5) |
|---|---|---|
| Sex | | |
| Male | 14 (56%) | 4 (80%) |
| Female | 11 (44%) | 1 (20%) |
| Age group (years) | | |
| < 5 | 0 (0%) | 0 (0%) |
| ≥ 5 - < 25 | 5 (20%) | 0 (0%) |
| ≥ 25 - < 39 | 14 (56%) | 4 (80%) |
| ≥39 | 6 (24%) | 1 (20%) |

**Table 5. Confirmed hepatitis A cases in Districts A and B by date of onset**

| Date of onset (m/d/yyyy) | No of cases |
|---|---|
| 1/1/2016 | 1 |
| 1/2/2016 | 1 |
| 1/3/2016 | 1 |
| 1/4/2016 | 4 |
| 1/5/2016 | 4 |
| 1/7/2016 | 5 |
| 1/8/2016 | 2 |
| 1/9/2016 | 1 |
| 1/11/2016 | 2 |
| 1/12/2016 | 4 |
| Total | 25 |

### Estimate the incubation period from the epidemic curve

From the outbreak investigation and the epidemic curve, we know that the first hepatitis A exposure took place in village A on 30 December 2015; three of the first cases of hepatitis A were cousins who visited each other on 30 December 2015. Therefore, the estimated incubation period range for the hepatitis A outbreak in Country X is at least two days (the incubation period is not a particular number of days but rather a range of days, reflecting differences in the intensity of exposures and/or differing immune responses among the exposed).

### Determine the possible source of transmission from the epidemic curve

There are serval types of epidemic curve. From the appearance of the epidemic curve (Fig. 6), this would seem to be a continuous source outbreak, i.e. all residents of Districts A and B (population at risk) are continuously exposed to the virus over an extended period of time. In a continuous source outbreak, there is no clear peak; the number of cases rises and plateaus rather than tapers off. While the plateau is not yet obvious in this outbreak, it is too early to dismiss a continuous source outbreak.

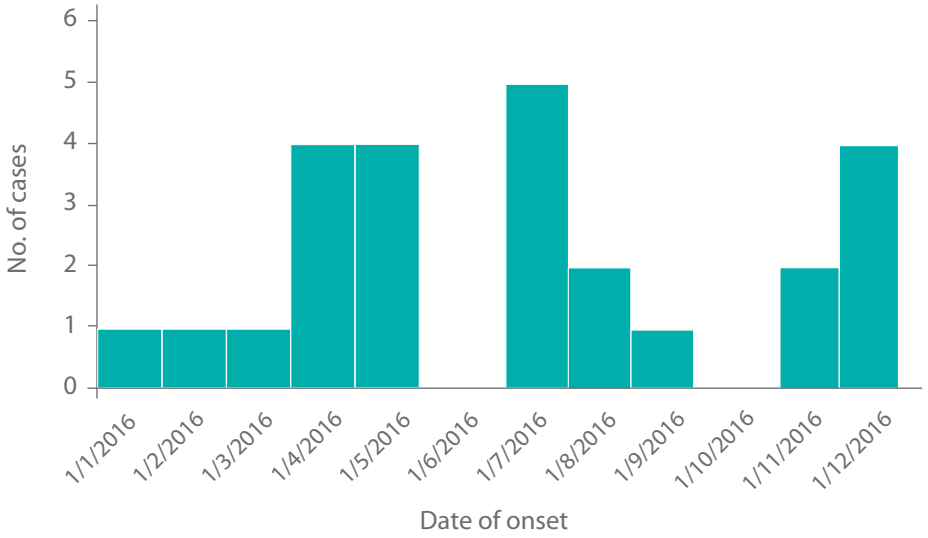



**Fig. 6. Epidemic curve of hepatitis A outbreak, January 2016**

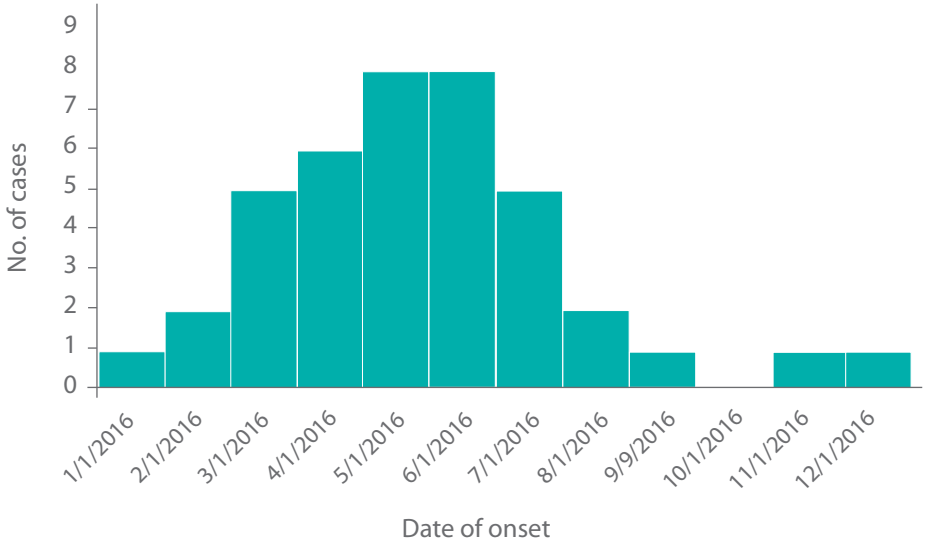



**Fig. 7 Example of a point source outbreak epidemic curve**

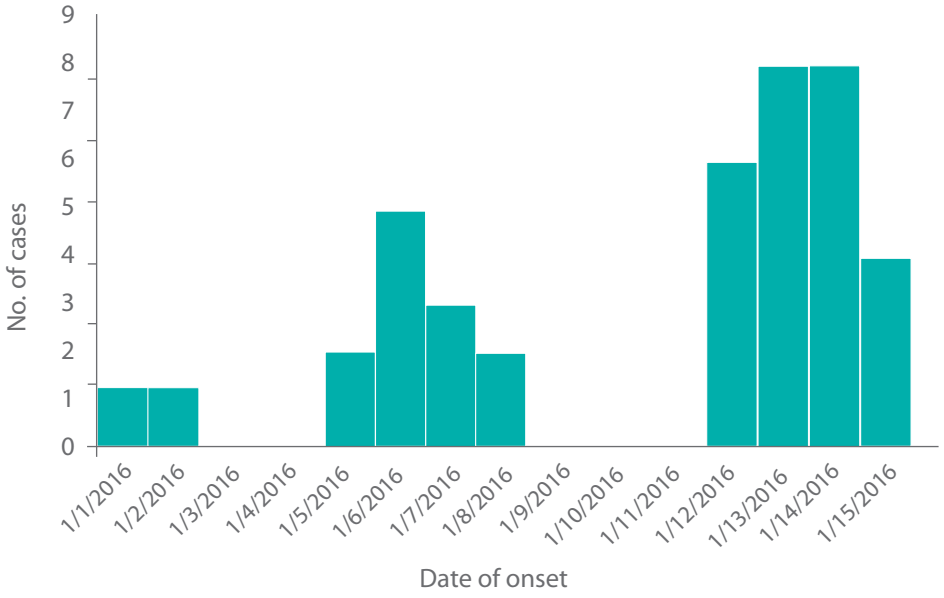



**Fig. 8. Person-to-person transmission epidemic curve**

Other epidemic curves suggest different outbreak types.

- Point source outbreak means that all the population at risk is exposed to the pathogen over a short period of time. In a point source outbreak, we expect the numbers to increase rapidly, reach a peak, and then gradually taper off. A typical point source outbreak epidemic curve is shown in Fig. 7.
- Person-to-person transmission outbreak. In this type of outbreak, the population at risk is infected through contact with members of the population that have been infected by the pathogen, including by those who are infected but are asymptomatic. The epidemic curve will appear to have multiple peaks as wave after wave of infection spreads through the population; cases in one peak are the source of infection for cases in subsequent peaks. A typical person-to-person transmission outbreak epidemic curve is shown in Fig. 8.

Based on what we know about the disease so far, the best formulated hypothesis is one that includes hepatitis A virus as the most likely agent of infection, spread through a common source. Removal from, or elimination of, the source of infection should constitute part of the interventions aimed at controlling the spread of the outbreak. These should include personal hygiene, improved water quality and sanitation facilities.

**Calculate the incidence and mortality for the outbreak**

Attack rate and case–fatality rate are the preferred measures of incidence and mortality respectively in an outbreak.

Attack rate: Number of cases of the disease/number of people at risk of infection

In our example:

Number of cases of hepatitis A in districts
A and B = 25

Number of people at risk of hepatitis A virus infection (residents of the 10 affected villages) = 600

Attack rate per 1000 population = (25/600) × 1000 = 41.7 cases of hepatitis A per 1000 population

Case fatality rate: Out of the 25 people with hepatitis A, 5 died; of the remaining 20 infected people, 5 are still ill and the outcome of their illness is unknown.

Therefore, the case–fatality rate currently = (5/25) × 100 = 20%

**Estimate the basic reproduction number and its implications for control of the hepatitis A outbreak**

At this point in the outbreak investigation, your primary concern is to contain the spread of hepatitis A.

The reproductive rate ($R_0$) is the number of secondary infections per 1 infected case. This measure is useful in determining the outbreak control measures. It is defined by the following equation:

$R_0 = \beta \times c \times D$

where:

β = average probability a contact will be infected over the duration of the relationship (this depends on the biology and behaviour of the infection)

c = average rate of coming into contact with an infectious case (isolation and quarantine minimize this parameter because infectious individuals and those who are suspected to be infected are in minimal contact with healthy individuals)

D = average duration of infectiousness (this parameter represents the length of symptomatic disease). We can modify D by shortening the duration of infectiousness. This could be achieved by proper treatment of cases, as well as improving the general health of the population at risk.

To reduce $R_0$, interventions must target one or more of the three parameters of $R_0$ (β, c and D) as shown below.

- Closing schools will alter c.
- Improving water quality will alter c.
- Precautionary measures at the hospitals (appropriate infection prevention and control procedures) will alter β.
- Risk communication messages about personal hygiene, water and sanitation will alter c.
- Early identification and treatment of symptoms will alter D.

## Summarizing data

After analysing EWARN data by person, time and place, irrespective of whether it is alerts data in the alert log, weekly reporting data in the database or line list data about an outbreak, it is useful to summarize what you know.

- Pick any aberration from the normal pattern and report or investigate for possible outbreak detection.
- Identify the start date of a possible outbreak and the end date when no more cases are detected to announce the beginning and end of an outbreak to relevant stakeholders, including health workers in the affected areas, various sectors participating in the response and decision-makers.
- Describe clustering by either residence or worksite or any other relationship in a specific time, and identify possible risk factors of occurrence that may be associated with a given place of occurrence.
- Describe the disease trend, either increase or decrease or stationary in time, to determine the evolution of the outbreak over time to determine if time or season (e.g. rainy season, dry season) is a factor influencing occurrence.
- Describe the affected population by age group, sex and occupation or other demographic characteristics to determine whether demographic characteristics influence occurrence.
- Describe the extent of disease spread by places where the highest number of cases and incidence rates were reported.
- Describe unfavourable outcomes, i.e. complications, disability and death rates.
- Describe the possible risk factors and source of infections.

The main purpose of data analysis is to generate actionable information from the epidemiological data. This purpose can only be served if the information generated is made available to all actors who contribute to the prevention and control of communicable diseases occurrence and outbreaks.

# 4. Data interpretation, feedback and information dissemination

## Data interpretation, feedback and information dissemination

Data interpretation translates analysis into key issues that are:

- easily understandable;
- actionable;
- targeted for specific stakeholder audiences.

Data interpretation clarifies why disease occurrence patterns are happening, and what that means for possible intervention recommendations.

## Interpretation

Once the analysis is completed, data must be interpreted and presented in a way that makes the key issues emerging from the data analysis easy to understand to stakeholders. These stakeholders include clinicians who provide care to infected people in health facilities, communities affected by the infectious disease, health professionals who address other aspects of prevention and control, such as health promotion, surveillance, laboratory support, as well as decision-makers who provide critical resources for prevention and control efforts.

The interpretation of the results of the analysis showing the distribution of the infectious disease by places affected, persons infected, and trend over time is where we attempt to address two key questions:

**"Why?" the observed distribution pattern from the analysis**

Specifically why:

- are cases distributed or occurring in certain place(s) and not others
- are certain age or gender groups affected more than others
- do cases appear to occur or increase at a certain time of the year or season?

This question should lead us to determine whether the clustering of cases around a certain community close to the river may be because of the shared use of contaminated water from the river; or older age groups may be affected more than the younger age groups because of waning immunity to a vaccine-preventable disease; or cases may be increasing during the rainy season because of increased contamination of water sources by run-off water; or because of increased proliferation of vectors.

**"What?" does it all mean in terms of prevention and control of the disease**

Knowing why cases are occurring in a given place, among individuals with certain demographic characteristics, and at certain times of the year or during a certain season leads us to possible environmental exposure risk factors, individual vulnerabilities such as low immunity, or risk factors associated with seasonal weather characteristics that favour the spread or transmission of the infectious agent of interest. If we have possible answers to the question of "WHY" disease occurrence patterns are happening, in terms of possible risk factors associated with the disease, we can respond to the question of "SO WHAT?" by attempting to identify and propose possible interventions to address the risk factors or drivers of the outbreak to stakeholders concerned about the outbreak or those involved in prevention and control

## Feedback and information dissemination

Most of the stakeholders may not have a health background or training in epidemiology. Therefore, the information from the interpretation of the findings must be presented in simplified language, and graphs should be used to convey complex data and trends in an easily understandable way. Clear recommendations should be made to the stakeholders directly involved in prevention and control. These recommendations should highlight priority areas and needs; for example, increasing numbers of diarrhoeal disease cases in a particular camp may highlight the need for special attention to water and sanitation in that area.

Furthermore, the epidemiological information should be presented in a user-friendly format, whether in the form of an epidemiological bulletin, a presentation or a media brief. In addition, mechanisms for disseminating the information (or feedback) to relevant stakeholders should be put in place. Depending on the magnitude and severity of an outbreak, and the information needs of the stakeholders involved in response and decision-making, information dissemination strategies may include: presentation of updates in sessions specially designated for this purpose; media briefings to reach out to the affected communities to avoid misinformation and undue anxiety; and sharing of epidemiological bulletins through electronic means, such as email and website posts, to those stakeholders who can be reached through these means.

# Annex: Establishing a disease threshold

## What is a threshold?

A threshold is a marker that alerts public health officials to take action. It uses past data to decide if an event is abnormal in order to identify possible outbreaks using surveillance data. Although there are internationally accepted signals that constitute an alert for some conditions, there are no globally fixed criteria for all diseases to guide outbreak detection.

## How to set disease thresholds

### Thresholds in exceptional circumstances

- Diseases under elimination, e.g. a single cases of poliomyelitis constitutes an alert.
- Emerging and re-emerging epidemic-prone diseases, e.g. a single case of Middle East respiratory syndrome coronavirus or Ebola constitutes an alert.
- Characteristics and settings of a population, such as overcrowded camp settings of displaced populations at risk following man-made or natural disasters. In such settings a single case of meningitis constitutes an alert.

### Thresholds for endemic disease

For high frequency, less severe infectious diseases, several cases are normally expected in endemic settings and a single case is not abnormal. However, we would want to know and investigate further when a disease occurs at levels exceeding expectation (or "baseline") by a significant amount. A threshold helps us decide if what we are seeing is abnormal and needs to be investigated further. Typhoid, for example, is a high-frequency disease in developing countries. Every reported case does not necessarily require a public health response. Only when cases are clustering within a defined area or there is a substantial increase in the occurrence of cases is follow-up required.

## Methods for calculating thresholds for endemic diseases

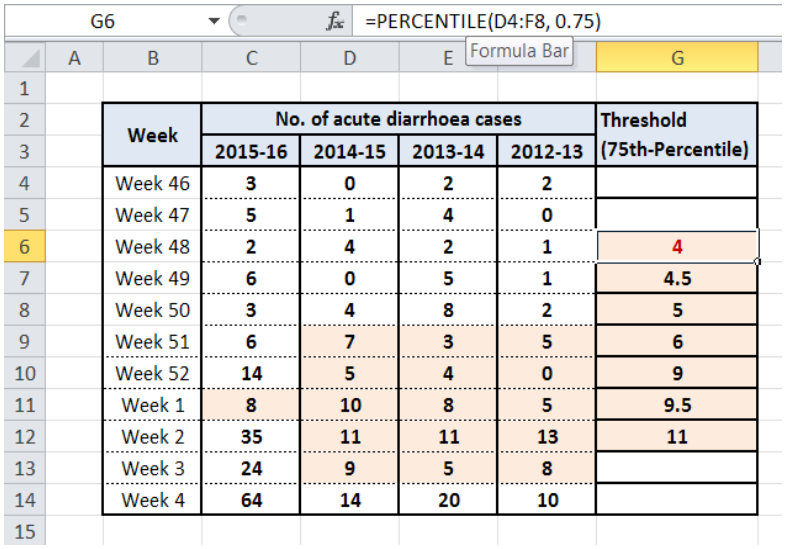
### Historical limits method

The historical limits method is more useful for diseases without any particular pattern or regular seasonality. It needs historical data for at least three years. If data are available for more than three years, they should be included. If population numbers are likely to fluctuate considerably depending on the population/area of interest, relative case counts (e.g. cases per 1000 person) rather than absolute counts should be used. If there was a major epidemic within the time that historical data are available (e.g. current year of interest 2018, major outbreak in 2016), consider removing the data for that year (2016 data). However, what would be considered an "epidemic year" needs to be carefully determined.

The historical limits method compares the cases reported in a given week of interest to two weeks before and two weeks after the same week in the past three years. Accordingly, we will have bandwidth of five weeks for the past three years, which gives 15 weeks.

This historical limits method has two alternatives approaches.

Centile approach: The threshold for early detection of outbreaks is calculated based on a decided level of centile, for example 75th centile. The centile level is decided by trial and error. Start by setting a low centile (e.g. 75%) and identify how many times it will be passed in a certain period of time and what the ratio of true outbreaks is among alerts detected using the 75% centile (**Figs A1 and A2**). Calculate the positive predictive value of the 75% level and decide on a higher threshold or keep it at that level.




| Week | No. of acute diarrhoea cases | | | | Threshold (75th-Percentile) |
|---|---|---|---|---|---|
| | 2015-16 | 2014-15 | 2013-14 | 2012-13 | |
| Week 46 | 3 | 0 | 2 | 2 | |
| Week 47 | 5 | 1 | 4 | 0 | |
| Week 48 | 2 | 4 | 2 | 1 | 4 |
| Week 49 | 6 | 0 | 5 | 1 | 4.5 |
| Week 50 | 3 | 4 | 8 | 2 | 5 |
| Week 51 | 6 | 7 | 3 | 5 | 6 |
| Week 52 | 14 | 5 | 4 | 0 | 9 |
| Week 1 | 8 | 10 | 8 | 5 | 9.5 |
| Week 2 | 35 | 11 | 11 | 13 | 11 |
| Week 3 | 24 | 9 | 5 | 8 | |
| Week 4 | 64 | 14 | 20 | 10 | |

**Fig. A1. Generation of thresholds based on the 75% centile**

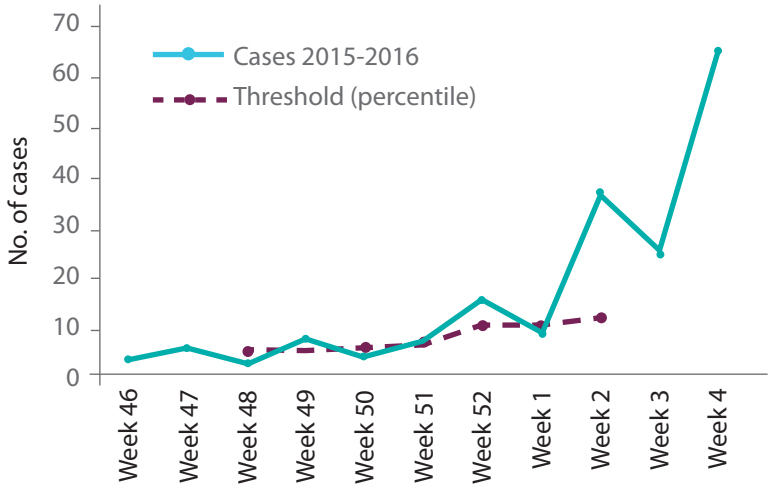



**Fig. A2. Trend of acute diarrhoea with a plotted threshold curve based on the historical limits method using centiles**

Mean and standard deviation approach: In the second approach, the threshold is generated by calculating the mean (i.e. using same time frame, 15 data points) + 1.96 × standard deviation. The generated threshold **(Fig. A3)** eliminates the "outbreaks" in week 49, 2015 and week 1, 2016 triggered using the centile method **(Fig A4)**, but clearly captures the epidemic in week 2, 2016.

I6 =G6+1.96*H6

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | |
| 2 | | Week | No. of acute diarrhoea cases | | | | Mean | sd | Threshold (mean+1.96*sd) |
| 3 | | | 2015-16 | 2014-15 | 2013-14 | 2012-13 | | | |
| 4 | | Week 46 | 3 | 0 | 2 | 2 | | | |
| 5 | | Week 47 | 5 | 1 | 4 | 0 | | | |
| 6 | | Week 48 | 2 | 4 | 2 | 1 | 2.4 | 2.2 | 6.8 |
| 7 | | Week 49 | 6 | 0 | 5 | 1 | 3.1 | 2.4 | 7.9 |
| 8 | | Week 50 | 3 | 4 | 8 | 2 | 3.4 | 2.4 | 8.1 |
| 9 | | Week 51 | 6 | 7 | 3 | 5 | 4.5 | 3.0 | 10.3 |
| 10 | | Week 52 | 14 | 5 | 4 | 0 | 6.4 | 3.7 | 13.7 |
| 11 | | Week 1 | 8 | 10 | 8 | 5 | 6.9 | 3.5 | 13.8 |
| 12 | | Week 2 | 35 | 11 | 11 | 13 | 8.9 | 4.8 | 18.3 |
| 13 | | Week 3 | 24 | 9 | 5 | 8 | | | |
| 14 | | Week 4 | 64 | 14 | 20 | 10 | | | |
| 15 | | | | | | | | | |

**Fig. A3. Generation of thresholds for acute diarrhoea based on the historical limits method using mean and standard deviation**

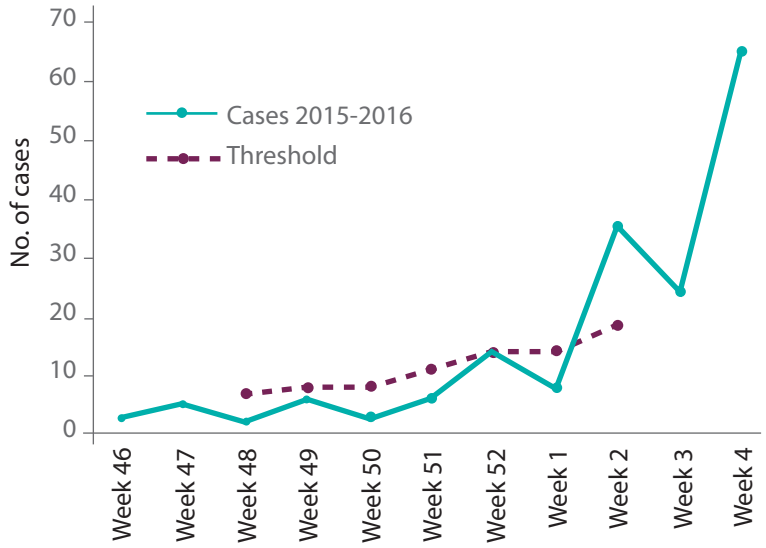



**Fig. A4. Trend of acute diarrhoea with plotted threshold curve based on the historical limits method using mean and standard deviation**

## Cumulative sum method

The cumulative sum (CUSUM) method needs minimal data, only number of cases by time in the current year. It should only be used with daily or weekly surveillance data, not monthly. In addition, using the average of previous weeks assumes that there is no strong seasonal variation in the endemic pattern of the communicable disease. The CUSUM method involves taking the average of the previous 7 data points with a 2-data point lag. Take each point of surveillance data and calculate the mean and standard deviation from a set of seven points in the past, with a lag of 2 points. You can use Excel to calculate the mean and standard deviation as illustrated below. Using surveillance data of the third week of 2016 as an example, use data from week 46 to week 52 of 2015 and ignore weeks 1 and 2 of 2016. Calculate the mean and standard deviation, then multiply the standard deviation by 2, and add it to the mean **(Fig. A5)**. Plot these points along the trend graph to get the threshold **(Fig. A6)**.

F12 =D12+E12*2

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 2 | | **Week** | **Cases 2015-16** | **Mean** | **Std-Dev** | **Threshold** |
| 3 | | Week 46 | **3** | | | |
| 4 | | Week 47 | **5** | | | |
| 5 | | Week 48 | **2** | | | |
| 6 | | Week 49 | **6** | | | |
| 7 | | Week 50 | **3** | | | |
| 8 | | Week 51 | **6** | | | |
| 9 | | Week 52 | **14** | | | |
| 10 | | Week 1 | **8** | | | |
| 11 | | Week 2 | **35** | | | |
| 12 | | Week 3 | **24** | 5.6 | 4.0 | **13.6** |
| 13 | | Week 4 | **64** | 6.3 | 3.9 | **14.2** |
| 14 | | | | | | |

**Fig. A5. Cumulative sum method to generate the threshold based on 7-week moving average**

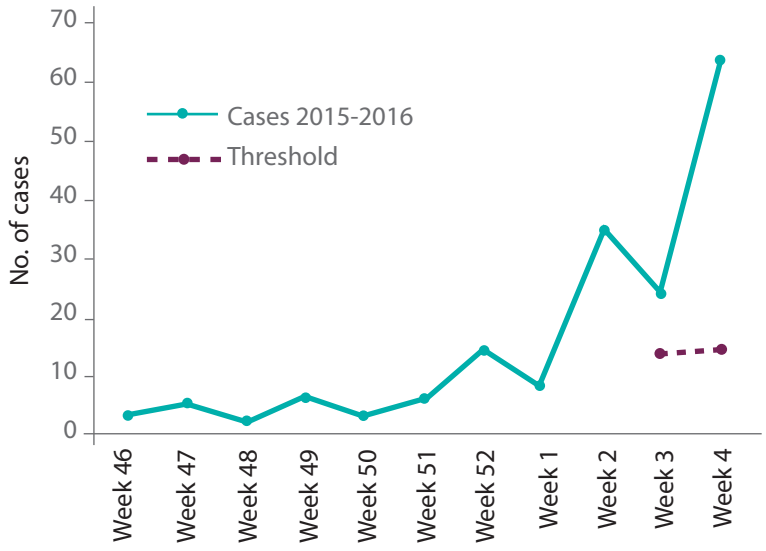



**Fig. A6. Trend of acute diarrhoea with plotted threshold curve based on the cumulative sum method**

# How to assess performance/modify the thresholds

Calculate the positive predictive value.

Positive predictive value = No. of outbreaks verified/no. of times the threshold was crossed × 100.

If the positive predictive value is too low, this means your threshold value is too low and you need to try the 80 or 85 centile and so on until you reach a suitable positive predictive value and threshold level for your local setting.

This handbook is intended for country-level epidemiologists and surveillance officers who work in data analysis and the production of epidemiological reports. It covers key concepts of epidemiological data analysis. Different types of surveillance data that are commonly generated by early warning surveillance systems are described and epidemiological indicators for analysis of early warning surveillance data are explained. A deliberate attempt has been made to simplify the often intimidating complexities of epidemiological data analysis. A clear step-by-step approach has been followed, and essential epidemiological indicators have been used to illustrate analyses of epidemiological data that are commonly encountered. Users of the handbook are shown how to perform sound data analysis, interpret findings and prepare epidemiological reports that are informative to policy-makers and stakeholders.

9 789292 745042