Data analysis and interpretation

9.1 Key points

> **Success**
>
> - A key goal for surveillance is to use data for public health action. Effective data analysis and reporting focuses on presenting data which can be interpreted in a standardized way in order to inform public health actions.
> - Data analysis is undertaken at the following points in the EWAR cycle:
> - Assessment of IBS {signals/alerts} against the alert thresholds
> - Calculating morbidity and mortality indicators for surveillance
> - Estimating outbreak indicators to monitor the outbreak
> - Monitoring the performance of the EWAR function.
> - Interpretation of surveillance data should take into account aspects related to person, time, and place.

9.2 Basics of analysis and reporting for EWAR

The value of EWAR data is weakened without rapid analysis and interpretation at the level that the data is collected.

Therefore, it is important to have a basic data analysis plan and human resources responsible for analysis and interpretation as the EWAR function is implemented.

While data analysis can be automated if electronic data collection packages like EWARS in a Box are used, interpretation is a human-led process.

Data analysis is typically related to weekly collection of IBS data, but is also relevant to types of EBS where scheduled reporting is used (e.g., community-based surveillance of {signals/alerts} during an outbreak).

Purpose of data analysis for EWAR

The analysis of collected data is a core step found in multiple places in the EWAR process:

- To characterize disease trends and aberrations that serve as the early detection of IBS signals

- To monitor the performance of the EWAR system (e.g., numbers of {signals/alerts} received, timeliness and completeness of reporting from sites, etc.)

- To produce surveillance and outbreak indicators for morbidity and mortality

- To evaluate the performance of the system

Box 1 outlines the guiding questions for IBS analysis.

Guiding questions for weekly IBS analysis

Are there {signals/alerts}?

- Have any priority diseases or other public health events of concern been detected during the reporting period (this week, for example)?

- Is an epidemic or unusual public health event suspected?

What are the weekly trends in person, time, and place?

- Of the cases, deaths or events detected, how many were confirmed?

- Where did they occur?

- How does the observed situation compare to previous observation periods of time this year? For example, when compared to the start of the reporting period, is the problem increasing?

- Are the trends stable, improving, or worsening?

- Is the reported surveillance information representative enough of the reporting site's catchment area? Out of all the sites that should report, what proportion has actually reported?

- How timely were the data received from the reporting sites? From: IDSR technical guidelines, 2nd edition.

Cleaning the data

A simple data cleaning process is a vital step to ensure accuracy and valid conclusions during the analysis. This should be done whether data is collated manually or automated through an

electronic tool. The following steps can be used as a checklist for cleaning practice for data officers:

- Use a unique ID and date system for reporting sites to systematize data reporting and enable retrospective review of entries.

- Check for late entries or missing entries by reporting sites. The current dataset could include the previous week's late entry and would need to be adjusted accordingly.

- Ensure that zero reporting practice is being followed. That is, check for the entry of a "0" where no cases were reported, or a blank space. Ensure that blank spaces mean that the reporter forgot to enter this data, or whether there were no cases.

- Check for double entries of weekly data.

- If there is doubt, verbally clarify data with reporting sites over the phone.

Population denominators

Population denominators are important for calculating rates of disease. The number of cases, those treated, and services delivered (e.g. the numerator) are not sufficient alone to evaluate the magnitude or needs in the population.

If available, the population at-risk (signified by the denominator) should be compared with the numerator in order to adequately monitor any changes over time against a potentially shifting population size. For the purposes of EWAR, denominators should be disaggregated only where it may affect determination or monitoring of an outbreak or public health emergency. For example, simple age groups (\< 5 years and 5 and over years) may be useful for diarrhea trends, while sex disaggregation is not as important to outbreak detection, at this early stage.

On the other hand, accurate population denominators are infrequently available in emergencies, in remote areas, and unstable situations. In such cases, it is useful to review potential recent sub-national data sources including census data for oral poliomyelitis campaigns, recent community health worker census data, emergency population estimation data, and satellite imagery estimation which may be available to support operations.

Frequency of data analysis

IBS surveillance data is usually analyzed on a weekly basis and compared with the results from previous week(s) to show the change over time. However, during outbreaks and emergencies,

data analysis should occur on a daily basis. Outside of an outbreak scenario, daily analysis is too frequent, and monthly analysis will not capture early warning rapidly enough. Data should be analyzed on a district (and/or other peripheral) level to investigate at a sufficiently detailed level.

9.3 EWAR indicators

The key indicators for routine surveillance using IBS data include incidence rate, attack rate, and case fatality ratio. See Table 1.

Table 1 Surveillance and outbreak indicators

Indicator	Numerator	Denominator	Rationale
Population under surveillance	--	Number of persons under surveillance in a given week	The population (if available) is important for calculating the incidence of disease. Population by age group (\<5, 5 +) and sex is valuable, if available.
Total reporting sites	--	Number of reporting sites in a given week	The number of sites gives an indication of the coverage of the system.
Number of cases* *suspect, probable, confirmed	# of new cases that occur within a given period of time		
(usually per week) in a given area	--	Weekly tallying of cases of disease. Number of cases should be presented as an epidemic curve to show time trends.	
Incidence rate	Number of new cases that occur within a given period of time (usually per week) in a given area = # cases in one week x1000 population	Population at risk*per 1,000, 10,000 persons at risk, or even more in case of small numbers.	To demonstrate a weekly trend and distribution of disease. Can be compared across population groups since it is adjusted for population size.Affected by uncertainty around the population size and affected by the level of reliance on healthcare.
Attack rate* (AR): cumulative incidence of disease over a given time period (e.g. one year,			
or the whole	Number of cases	Population	AR indicates the

9.4 Key EWAR performance indicators

Consistent monitoring of system performance on a weekly basis is key for assuring that the EWAR function is working effectively.

Producing performance indicators by reporting site is helpful for planners and the reporting sites to monitor and improve their capacity for early warning. The following basic indicators should be included in a weekly report, alongside the epidemiological indicators above (Table 2).

Table 2 Performance indicators

Domain	Indicator	Numerator	Denominator	Rationale
Completeness of reporting	Number of reports submitted in a given week	Number of reports submitted in a given week	Number of reporting sites	Weekly tally and comparison of the number of reports submitted. The proportion of sites submitting reports in a given week should not drop below 80%.
Timeliness of reporting	Proportion of total reporting sites submitting a report in a given week by the stated deadline	Number of reports submitted in a given week by the stated deadline	Number of reports expected	Weekly tally and comparison of the number of reports submitted on time. The proportion of sites submitting reports in a given week and on time should not drop below 80%.
Signal management	Proportion of {signals/alerts} detected and verified	Number of new {signals/alerts} detected and verified	Number of new {signals/alerts}	Monitor the effectiveness of the verification function.
{Signal/alert} management	Proportion of {signals/alerts} detected and verified within 24 hours (verified)	Number of new {signals/alerts} detected and verified within 24 hours	Number of new {signals/alerts} detected	Monitor the effectiveness of the verification function.
{Signal/alert} management	Proportion of events assessed within 48 hours	Number of new events risk assessed within 48 hours	Number of new events detected	Monitor the effectiveness of the response function.

9.5 Analysis plans

Data should be analyzed in a simple way using a predictable approach.

IDSR guidelines provide an outline of a data analysis plan which can support the EWAR function of weekly IBS surveillance (Box 2).

> **Weekly analysis plan (sample)**
>
> 1. Calculate completeness and timeliness of reporting Monitoring whether surveillance reports are received on time and if all reporting sites have reported is an essential first step in the routine analysis of the surveillance system. This assists the district (or other level) surveillance team in identifying silent areas (areas where health events may be occurring but which are not being reported) or reporting sites that need assistance in transmitting their reports.
>
> 2. Calculate district (or other level) totals by week (or by month). Update the total number of reported cases and deaths for the whole year. This is summary information that helps to describe what has happened in the particular reporting period.
>
> 3. Prepare cumulative totals of cases, deaths and case fatality rates since the beginning of the reporting period.
>
> 4. Use geographic variables (such as hospitals, residence, reporting site, neighborhoods, village and so on) to analyze the distribution of cases by geographic location. This is information that will help to identify high risk areas.
>
> 5. Analyze disease trends for at least the diseases of highest priority in your district. Compare trends against alert thresholds. Monitor the trends for cases, deaths, and case fatality rates to identify any unusual increases or disease patterns.

9.6 Good data reporting practice

Presenting a summary of the IBS analysis on a weekly basis is a key step in communicating important findings to a wider audience and giving feedback to reporting sites.

The key elements and analyses to be presented are shown in Table 3.

Table 3 Key elements for epidemiological bulletins including EWAR

Element	Analysis
Burden of disease	- A summary analysis showing numbers of cases and deaths by disease, and by week, and compared to cumulative (and/or previous week) - Graphs of proportional morbidity
Time	Graphs of incidence, by disease (epidemiologic curve)
EWAR component	- A tabulation of IBS and EBS {signals/alerts} - Proportion {signals/alerts} verified - Investigation and response actions taken
Surveillance system and performance	- The population under surveillance (size, geographic area) - Completeness of reporting (by site or district)- Timeliness of reporting (by site or district)
Interpretation	Focus on the epidemiologicalsituation and comparison ofthe current week compared to the previous or cumulative period.