GLOBAL TASK FORCE ON CHOLERA CONTROL

Identification of PAMIs for cholera control

Transcript of online course

MODULE 2

Data preparation

Slide 1

Welcome to Module 2 of the GTFCC online course on the identification of PAMIs for cholera control.

Slide 2

This module focuses on the compilation and the preparation of the data to identify PAMIs for cholera control.

Slide 3

After completing this module, you will be able to:

- Explain how to set the PAMI analysis period and the geographic level of PAMIs;
- Describe the data to be compiled to identify PAMIs for cholera control;
- Explain the importance of data cleaning;
- Explain how to handle missing data;
- Use the GTFCC data model template.

Slide 4

Let's look into the data that need to be compiled to identify PAMIs for cholera control.

Slide 5

Prior to actual data compilation, the time period over which data will be compiled and the geographic level of data compilation must be determined.

The time period over which data will be compiled corresponds to the PAMI analysis period.

The geographic scale at which data will be compiled corresponds to the geographic level of PAMIs.

Both the PAMI analysis period and the geographic level of PAMIs should be set in agreement between multisectoral stakeholders who will be involved in PAMI identification or in NCP development or implementation.

Slide 6

As a general principle, the PAMI analysis period should be at least five years. Longer PAMI analysis period may also be considered, up to 15 years.

To determine the duration of the PAMI analysis period, the following should be considered.

First, the availability of retrospective surveillance data considering both epidemiological data and data on cholera tests.

The quality and the comparability of surveillance data over time should also be considered in order to have, as much as possible, a coherent and consistent dataset.

The historical patterns of cholera outbreaks in the country should also be considered, in particular the usual duration of inter-epidemic periods - as relevant in the country considered.

The year of occurrence of any events that may have led to a major increase or a major decrease in the cholera risk should also be considered as relevant in the country-specific context.

Slide 7

Let's illustrate the selection of the PAMI analysis period with a fictive example.

In this example, a country initiates the identification of PAMIs for cholera control in 2024, and assess the availability, the quality, and the comparability of cholera surveillance data over time in order to decide on its PAMI analysis period.

This assessment shows that the completeness of cholera reporting decreased in 2020 in the context of the Covid-19 outbreak, and resumed in 2021.

In addition, in this country, the cholera surveillance strategy changed in 2016 with updated case definitions and revised testing strategies. These changes in the surveillance strategy resulted in improved specificity of cholera surveillance.

In this country, it would be advisable to consider a seven-year PAMI analysis period, from 2017 to 2023.

Indeed, a PAMI analysis period longer than five years would mitigate the lack of reliability of the data in 2020.

On the other hand, it is best not to include in the analysis data prior to 2017 considering the major change in the surveillance strategy that would compromise the comparability of the data across the analysis period.

Slide 8

Moving on to the selection of the geographic level of PAMIs. Determining the appropriate geographic level of PAMIs is always a country-specific decision to be made by country stakeholders.

The following should be considered to determine the appropriate geographic level of PAMIs:

- If cholera surveillance data is aggregated, the lowest level at which retrospective surveillance data is available over the PAMI analysis period should be considered.
- In addition, operational considerations regarding the implementation of multisectoral interventions in PAMIs should also be carefully considered.

If the geographic level of PAMIs is small, the resulting NCP may be overly fragmented and coordinating its implementation may be challenging.

On the other hand, if the geographic level of PAMIs is big, the resulting NCP may be overly demanding on resources to implement multisectoral interventions.

Experience shows that countries often select administrative level two or administrative level 3 as the geographic level of PAMIs.

The key to determine the appropriate geographic level of PAMIs is to anticipate the best operational balance for the NCP in the country specific context.

Slide 9

Once the PAMI analysis period and the geographic level of PAMIs have been determined, data is compiled for each geographic unit over the selected analysis period.

As a general principle, the identification of PAMIs does not require to generate new data. Instead, it is mostly about compiling, assessing, and cleaning existing data.

The data to be compiled include at a minimum retrospective surveillance data, covering both epidemiological data and data on test results, geographic data, and population data. In addition, and optionally as relevant in the country specific context, data on the presence or absence of vulnerability factors may also be compiled.

Because data from multiple sources have to be compiled for PAMI identification, data compilation requires close coordination, communication and collaboration across multiple sectors.

Slide 10

The epidemiological data to be compiled for each year of the PAMI analysis period and for each geographic unit of the country are the number of cholera cases, the number of cholera deaths, and the number of weeks with at least one reported cholera case.

Cholera cases and cholera deaths include both suspected cholera cases and cases tested positive for cholera.

Slide 11

The data on cholera tests to be compiled for each year of the PAMI analysis period and for each geographic unit of the country are the number of weeks with at least one suspected cholera case tested for cholera, the number of suspected cholera cases tested for cholera, and the number of suspected cholera cases tested positive for cholera.

For the purpose of PAMI identification, all cholera test methods are considered without distinction. That is rapid diagnostic tests, culture, or PCR.

Slide 12

Lastly, if vulnerability factors are considered in the PAMI analysis, recent data on the presence or absence of each vulnerability factor should be compiled at the geographic unit level.

For the PAMI analysis, data on the presence or absence of vulnerability factors are only required for the geographic units where the priority index may lack reliability. For the purpose of PAMI identification, compilation of data on vulnerability factors may focus on these geographic units.

However, data on the presence or absence of vulnerability factors may also be helpful following PAMI identification in particular to design strategies for preventive OCV. Therefore, out of convenience for the next steps, countries may decide to compile data on the presence or absence of vulnerability factors for all geographic units of the country.

Slide 13

Following the compilation of the data for the PAMI analysis, an essential step is to clean the data prior to their analysis.

Slide 14

Data cleaning is essential for the PAMI analysis to be reliable. Without proper data cleaning, incorrect conclusions may be drawn.

All data compiled for PAMI identification should be cleaned by an experienced data manager or data analyst following best practices in this field.

In particular, attention should be given to any duplicates in geographic units and any inconsistencies across variables.

Slide 15

Here are examples of inconsistencies that should be looked for in the dataset, and corrected as needed.

On one year, in a given geographic unit, it would be inconsistent to have cholera deaths if there were no cholera cases. Deaths is a potential outcome for a cholera case. In other words, cholera deaths are among cholera cases.

Following the same principle, a number of cholera deaths greater than the number of cholera cases would also be inconsistent.

As another example of potential inconsistency, persistence equal to zero would be inconsistent if cholera cases were reported. Likewise, if no cholera cases were reported, persistence different to zero would also be inconsistent.

As a last example, a number of cases tested positive greater than the number of cases tested would also be inconsistent.

Slide 16

Any errors or inconsistency detected in surveillance data at the data cleaning step should be flagged to surveillance officers or laboratory officers to be corrected in a reliable manner.

To correct the data, the surveillance or laboratory officers may go back to the original database at the central or at the local level. As needed, they may also consult historical situation reports or retrieve historical records at the reporting sites or at laboratories.

Cleaning surveillance data for PAMI identification and making retrospective corrections in surveillance data is an insightful process to identify any improvements to be made in record keeping, information systems, and data management for cholera surveillance. This should be built on when designing strategies to improve cholera surveillance as part of the NCP.

Slide 17

Essential to a reliable PAMI identification is also the management of missing data. Addressing missing data is essential to limit bias in the PAMI analysis.

Missing data should be clearly recorded in the dataset and, importantly, should be differentiated from zero values.

Zero values are for when the value of a variable is equal to zero.

Missing values are for when there is no information regarding the value of a variable due to missing data.

If missing data are not properly recorded, missing data cannot be addressed.

To address missing data, as a first step, missing surveillance data should be flagged to surveillance officers or laboratory officers for the missing data to be retrieved - similar to how inconsistencies in the dataset are corrected.

If despite efforts to retrieve missing data, some data cannot be retrieved, then how to handle missing data and limit bias in the PAMI analysis depends on the extent of missing data.

Slide 18

If there are missing data for one year, several geographic units and more than one indicator, the recommended approach is to exclude the corresponding year from the PAMI analysis.

If there are missing data for one year, several geographic units and one indicator, the recommended approach is to exclude the corresponding year from the calculations for this indicator.

If there are missing data for one year and a few geographic units, the recommended approach is not to rely on the priority index for the corresponding geographic units and, instead, to discuss the potential PAMI status of these geographic units at the stakeholder validation.

Lastly, if there are missing data for several years, the availability of any other data sources should be explored. If no sufficiently comprehensive data source can be identified, the recommended approach is to reconsider the PAMI analysis period.

Addressing missing data as part of PAMI identification is an insightful process to identify any improvements to be made to improve the completeness of reporting and recording. This should be built on when designing strategies to improve cholera surveillance as part of the NCP.

Slide 19

Once all necessary data have been compiled and have been cleaned, the next step is to format the data for the PAMI analysis.

Slide 20

The dataset should be formatted in accordance with the PAMI data model template.

A PAMI Excel tool is available to automatize all PAMI calculations. However, this tool can only perform the calculations if the PAMI dataset is formatted following the PAMI data model template.

The PAMI data model template is an Excel file which can be downloaded at https://tinyurl.com/PAMIcontrol or by scanning this QR code.

Slide 21

Here is how the PAMI data model template is structured. Each row is a geographic unit. Each column is a variable for PAMI identification.

In light grey are variables to identify the geographic units.

In dark grey are variables on the population for each year of the PAMI analysis period.

In green are variables on the number of cholera cases for each year of the PAMI analysis period.

In pink are variables on the number of cholera deaths for each year of the PAMI analysis period.

In light blue are variables on the number of cholera cases tested for cholera for each year of the PAMI analysis period.

In dark blue are variables on the number of cholera cases tested positive for cholera for each year of the PAMI analysis period.

In dark green is a variable on the total number of weeks in the PAMI analysis period.

In light orange are variables on the number of weeks with cholera cases reported for each year of the PAMI analysis period.

In dark orange are variables on the number of weeks with suspected cholera cases tested positive for cholera on each year of the PAMI analysis period.

Lastly, in yellow are optional variables on the presence or absence of vulnerability factors.

Very importantly do not modify the headings of columns, follow strictly the headings of the data model template. The only change that can be made in the heading of columns is the year number in order to reflect the selected PAMI analysis period.

Slide 22

By default, the PAMI data model template is for a five-year PAMI analysis period.

If your PAMI analysis period covers five years, you simply have to edit the year number in the heading of columns to match your PAMI analysis period.

If your PAMI analysis period is longer than five years, you need to add columns. For each annual variable, add as many columns as needed to match the duration of your PAMI analysis period. Make sure to follow the template to label the variable in each new column.

For example, in the PAMI Data model template, the variable on the number of cases must be labeled c_y_year number. For each additional year of the PAMI analysis period, a new column is added and is labelled following exactly this nomenclature.

Slide 23

As we wrap up this module, here are the important points to remember.

To prepare the data to identify PAMIs for cholera control, stakeholders should first determine the most appropriate analysis period and geographic level of PAMIs in their country-specific context.

Then, retrospective surveillance data, including both epidemiological data and data on cholera tests are compiled for each geographic unit and for each year of the PAMI analysis period.

Prior to analyzing the data, the dataset is carefully cleaned, and missing data are addressed.

The dataset is then formatted in accordance with the PAMI data model template so that calculations can be automated in the PAMI Excel tool.

Slide 24

Before moving on to the next module, we encourage you to take a short quiz. There are three questions in this quiz.

Slide 25

Question 1. What could be a potential issue if geographic units for PAMIs are defined at a very "small" geographic level?

a) The NCP may be too broad, and its implementation may be demanding on resources;

b) The NCP may be overly fragmented, and its implementation may be difficult to coordinate;

c) The compilation of retrospective epidemiological data may be unreliable;

d) Significant progress towards achieving cholera control may be too slow.

Slide 26

The correct answer is b. If geographic units for PAMIs are defined at a very "small" geographic level, the NCP may be overly fragmented, and its implementation may be difficult to coordinate.

Slide 27

Question 2. What is the first step if a missing epidemiological data is identified in the PAMI dataset?

a) It should be filled as zero.

b) It should be ignored during the analysis.

c) It should be marked as an outbreak as a precautionary measure.

d) It should be flagged to surveillance officers for retrieval.

Slide 28

The correct answer is d. The first step if a missing epidemiological data is identified in the PAMI dataset is to flag it to a surveillance officer to attempt to retrieve it.

Slide 29

Question 3. This is the last question. How to best describe the PAMI data model template?

a) It is a customizable template to be adapted to match countries' data structure.

b) It can only be used if the PAMI analysis period is 5 years.

c) It must be strictly followed for the data to be analyzed in the PAMI Excel tool.

d) Each column is a geographic unit.

Slide 30

The correct answer is c. The PAMI data model template must be strictly followed for the data to be analyzed in the PAMI Excel tool.

Slide 31

We have now completed this module.