

Identification of PAMIs for cholera elimination

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 elimination.

Slide 2

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

Slide 3

After completing this module, you will be able to:

- Explain how to determine the geographic level of PAMIs;
- Describe the data to be compiled to identify PAMIs for elimination;
- Explain how to handle missing data;
- And, use the GTFCC data model template.

Slide 4

Let's look at the data that need to be compiled to identify PAMIs for elimination.

Slide 5

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

The data to be compiled include: geographic data; population data; retrospective surveillance data as well as data on the presence or absence of vulnerability factors.

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

Slide 6

Prior to actual data compilation, the geographic level of PAMIs must first be determined. The data will then be compiled at this geographic level.

The geographic level of PAMIs is country-specific. It should be determined in agreement between multisectoral stakeholders who will be involved in PAMI identification or in NCP development or NCP implementation.

Slide 7

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

- If cholera surveillance data is aggregated, the lowest level at which retrospective surveillance data is available;
- In addition, and importantly, operational considerations regarding the implementation of multisectoral interventions in PAMIs.

If the geographic level of PAMIs is small, the 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 NCP may be overly demanding on resources to implement multisectoral interventions in PAMIs.

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 is to anticipate the best operational balance for the NCP in the country-specific context.

Slide 8

Moving on to the time period over which data should be compiled.

Retrospective surveillance data are compiled for at least the past 5 years. The occurrence of confirmed cholera outbreaks and the occurrence of confirmed imported cholera cases are documented in each geographic unit over this time period.

Data on the presence or absence of vulnerability factors in each geographic unit are compiled for the most recent year depending on the most up-to-date data source that is available for each vulnerability factor. This might correspond to a different time period for different vulnerability factors depending on data availability.

Slide 9

The vulnerability factors relevant in the country-specific context are identified in consultation between stakeholders representing multiple sectors.

As a starting point, the GTFCC indicative list of generic vulnerability factors should be reviewed in order to assess if any of these factors is not applicable or relevant in the country-specific context. In addition, whether any additional vulnerability factor, not included in the indicative list, is of particular relevance in the country-specific context is also assessed.

If you are unsure about what the indicative list of generic vulnerability factors is, we encourage you to go back to module 1 of this course.

Additional vulnerability factors not included in the indicative list are relevant in the country-specific context if they are demonstrated to increase the risk of introduction of cholera, onset of a cholera outbreak, or spread of a cholera outbreak.

Additional vulnerability factors can be identified with expert consultations or scoping reviews.

Slide 10

Once the list of vulnerability factors to be considered in the PAMI analysis has been established, the next step is to define a measurable vulnerability indicator for each vulnerability factor in order to assess its presence or absence at the geographic unit level. Each measurable vulnerability indicator should be associated with a data source.

Measurable indicators are essential for vulnerability factors to be interpreted in a reproducible manner and assessed in an objective manner.

Identifying a reliable data source for each measurable indicator is important for a data driven and reliable measurement of the presence or absence of each vulnerability factor.

Slide 11

There may not be a perfect data source for each vulnerability factor.

Compromise may have to be found to identify the best data source for each vulnerability factor taking into account and balancing the following: the date of the data to ensure it is reasonably recent; the reliability of the data; the availability of the data at the required geographic level; the completeness of the data in particular with regard to geographic coverage.

Some adjustments may have to be considered depending on the data available.

The definition of measurable vulnerability indicators may have to be tweaked to be measurable using available data sources. For example, to match existing categories or thresholds.

It may also happen that for some indicators no data is available at the geographic level of PAMIs but only at the upper geographic level. For example, if the geographic level of PAMIs is administrative level 3, it may happen that, for some vulnerability factors, data is only available at the administrative level 2. The recommended approach in this situation is for lower level geographic units to inherit the value of the upper level units regarding the presence or the absence of the factor considered.

Slide 12

To illustrate this, let's look into examples of measurable vulnerability indicators and associated data sources. These are fictive examples; relevant measurable vulnerability indicators and data sources must of course be customized in each country.

The vulnerability factor “high population density” may be defined differently in different countries. For example, it may be defined as a geographic unit with a population density above 1 000 inhabitants per square kilometer. Recent population figures may be available from the Ministry of Demography.

The vulnerability factor “unimproved water” may be defined as a geographic unit where above 30% of the population use unimproved service level or surface water or where above 15% of the population use surface water. The corresponding figures may be retrieved from the WHO UNICEF joint monitoring programme, the JMP. These data are usually not updated every year.

Slide 13

Once relevant vulnerability factors have been selected and a measurable indicator associated with a data source has been identified for each vulnerability factor, then the presence or absence of each vulnerability factor in each geographic unit can be assessed.

Importantly, if there is a gap in a data source and there is no information on the presence or absence of a vulnerability factor in a specific geographic unit, this should be recorded as missing. That way, the absence of data to assess the presence or absence of the vulnerability factor is clearly differentiated from the absence of the vulnerability factor.

Slide 14

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

Slide 15

Data cleaning is critical so that the PAMI analysis is 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.

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

Slide 16

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.

Importantly, all missing data should be filled before any decision is made on the list of PAMIs.

To that end, different strategies apply depending on the extent of data missingness, namely if there is partial data missingness or substantial data missingness.

Slide 17

There is substantial data missingness if data to assess the presence or absence of a vulnerability factor is missing for most geographic units. In that situation, it is recommended to look for an alternative data source or to revisit the definition of the measurable vulnerability indicator.

There is partial data missingness if data to assess the presence or the absence of a vulnerability factor is missing for only a few geographic units. In that situation, a survey might be considered to collect the missing data. A qualitative assessment based on expert opinions can also be performed.

Lastly, if under exceptional circumstances, it was not possible to fill some missing data with a survey or based on expert opinions, the remaining missing data is filled at the stakeholder validation based on participants' knowledge.

Slide 18

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

Slide 19

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

This is because a PAMI Excel tool is available to automatizes all PAMI calculations, however this tool can only perform the calculations if the dataset is formatted in accordance with the PAMI data model template.

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

Slide 20

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 is a variable on population.

In blue is a variable on the occurrence of confirmed cholera outbreaks.

In yellow are all variables on the presence or absence of vulnerability factors.

Variables labelled from VF2 to VF14 correspond to the generic vulnerability factors of the GTFCC indicative list.

Variables labelled from VF15 to VF18 are placeholders for any additional vulnerability factor considered relevant in the country-specific context.

Slide 21

When filling the data model template with your data, importantly, do not modify the headings of columns. In other words, no change should be made in variables names.

If additional country-specific vulnerability factors are considered in your PAMI analysis, fill the column labelled VF_15 to VF_18 without any modification to the heading of these columns.

Slide 22

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

To prepare the data to identify PAMIs for cholera elimination, stakeholders first determine the most appropriate geographic level of PAMIs in the country-specific context.

Retrospective surveillance data are compiled at the selected geographic level for at least the past five years.

Vulnerability factors relevant in the country-specific context are identified together with a measurable vulnerability indicator and an associated data source to assess the presence or absence of each factor at the selected geographic level.

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

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 23

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

Slide 24

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 data may be too cumbersome to expedite the data collection process.
- d) Significant progress towards achieving cholera elimination may be too slow.

Slide 25

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, as a result, its implementation may be difficult to coordinate.

Slide 26

Question 2. Why is it essential to fill all missing data?

- a) To ensure all geographic units have the same vulnerability index.
- b) To remove any vulnerability factor with missing data from the analysis.
- c) To avoid bias in the calculation of the vulnerability index.
- d) To verify that no data source with incomplete data coverage was used.

Slide 27

The correct answer is c. All missing data should be filled in order to avoid bias in the calculation of the vulnerability index.

Slide 28

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 vulnerability factors considered are those of the indicative list of generic vulnerability factors.
- 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 29

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 30

We have now completed this module.