PAMIs for cholera control

Module 2

GLOBAL TASK FORCE ON
CHOLERA CONTROL

PAMIs for cholera control

Data preparation

What will you learn?

- How to set the PAMI **analysis period** & the **geographic level of PAMIs**
- Which **data to compile** to identify PAMIs for control
- Why **data cleaning** is essential
- How to **handle missing data**
- How to use the **data model template**

Data compilation

Photo adapted from: WHO / Fabeha Monir

Data compilation – Time & Space

Prior to compiling any data to identify PAMIs, it is necessary to define:

- **The PAMI analysis period**

  This is the time period over which data will be compiled

- **The geographic level of PAMIs**

  This is the scale at which data will be compiled

The PAMI analysis period and the geographic level of PAMIs are defined in agreement between stakeholders across multiple sectors

PAMI analysis period

As a general principle, it is recommended to consider
at least the last 5 years (up to the last 15 years)

- The following is considered to set the analysis period:
  - **Availability** of surveillance data (epidemiological data & data on tests)
  - **Quality & comparability** of surveillance data over time
  - Historical patterns of cholera outbreaks and **duration of "inter-epidemic" periods**
  - Any event(s) resulting in a **major change in the cholera risk** (increase or decrease)

PAMI analysis period

Illustration

Scenario

A country initiates the identification of PAMIs for cholera control in 2024

- **Completeness of cholera reporting**
  - Decreased in 2020
  - Resumed in 2021
- **Cholera surveillance strategy changed in 2016**
  - Improved specificity

PAMI analysis period

A 7-year analysis period (**2017-2023**) would be advisable

- Selecting an analysis period > 5 years mitigates the lack of reliability of the data in 2020
- Not selecting an analysis period > 7 years avoids including data prior to 2017 that would not be comparable

Geographic level of PAMIs

The geographic level of PAMIs is defined in agreement between country stakeholders

- The following is considered:
  - **Availability of surveillance data**
    (if only aggregate data is available)
  - **Operational considerations**
    - "**Small**" geo units as PAMIs may result in an overly **fragmented** NCP
    - "**Big**" geo units as PAMIs may result in an overly **demanding** NCP

In the past, countries have often selected:

- **Administrative Level 2 ("districts", "counties")**

or

- **Administrative Level 3 ("municipalities")**

Data compilation

The identification of PAMIs does not require to generate new data but to compile, assess, and clean existing data from multiple sources

- Data to be compiled include:
  - Retrospective **surveillance data** (epidemiological data & data on tests)
  - **Geographic data**
  - **Population data**
  - And, optionally, data on the presence/absence of vulnerability factors

Data compilation requires coordination, communication and **collaboration across multiple sectors**

Epidemiological data

For each year of the analysis period and for each geo unit of the country

Number of **cholera cases***

Number of **cholera deaths***

Number of **weeks with ≥ 1 cholera case***

*Suspected or tested positive

Data on tests

For each year of the analysis period and for each geo unit of the country

Number of **weeks with ≥ 1 suspected cholera case tested*** for cholera

Number of **suspected cholera cases tested*** for cholera

Number of **suspected cholera cases tested positive*** for cholera

* Regardless of the testing method - rapid diagnostic test, culture, PCR

Data on vulnerability factors

Optional

Recent data on the presence/absence of vulnerability factors at the geo unit level

For PAMIs identification

- Only required for the geo units where the priority index is anticipated to lack reliability
- Countries may decide to focus data compilation on these **specific geo units**

Following PAMI identification

- Data on vulnerability factors will also be helpful (e.g., strategies for preventive OCV)
- Countries may decide to expand data compilation to **all geo units** in anticipation

Data cleaning

Photo adapted from: WHO / Christopher Black

Data cleaning

Data cleaning is essential for reliable PAMI identification
Without proper data cleaning, incorrect conclusions may be drawn

The dataset should be cleaned by an
experienced data manager or data analyst

- The following should be paid attention to:
  - **Duplicates** in geo units
  - **Inconsistencies** across variables

Examples of inconsistencies

- **On one year in a given geo unit**:

  - There cannot be cholera deaths if there were no cholera cases
  - Number of cholera deaths cannot be > number of cholera cases
  - Persistence cannot be zero if cholera cases were reported (and vice versa)
  - Number of cholera cases tested positive cannot be > number of cholera cases tested

Corrections

Errors & inconsistencies should be flagged to surveillance officer(s) and/or laboratory officer(s) for correction

- Correcting the data may require to:
  - Go back to the **original database/datafile** at the central or local level
  - Consult **historical situation reports**
  - Retrieve **historical records** at reporting sites or laboratories

Data cleaning is helpful to **identify improvements to be made** in record keeping, information systems, data management

Build on this when designing interventions in the NCP to improve cholera surveillance

Missing data

Recording and addressing missing data are essential to limit bias

Record missing data

In the dataset, differentiate **zero values and missing values**

- **Zero** is for when the value of the variable is zero
- **Missing value** is for when there is no information on the value of a variable

Address missing data

Step 1. Flag missing data to relevant officer(s) and/or laboratory officer(s) for **retrieval**

Step 2. If some data cannot be retrieved, how to handle missing data depends on the extent of missing data

Remaining missing data

Data missing for	Recommended strategy
1 year, many geo units, ≥1 indicator	Exclude this year from the analysis
1 year, many geo units, 1 indicator	Exclude this year from the calculations for this indicator
1 year, a few geo units	Assess these geo units at the stakeholder validation to determine if they should be additional PAMIs
Several years	Explore other data sources If no data source is identified, reconsider the analysis period

Build on this to design interventions in the NCP to improve the completeness of reporting/recording

Data formatting

Photo adapted from: WHO / Christopher Black

PAMI data model template

- A **PAMI Excel tool** automatizes all calculations

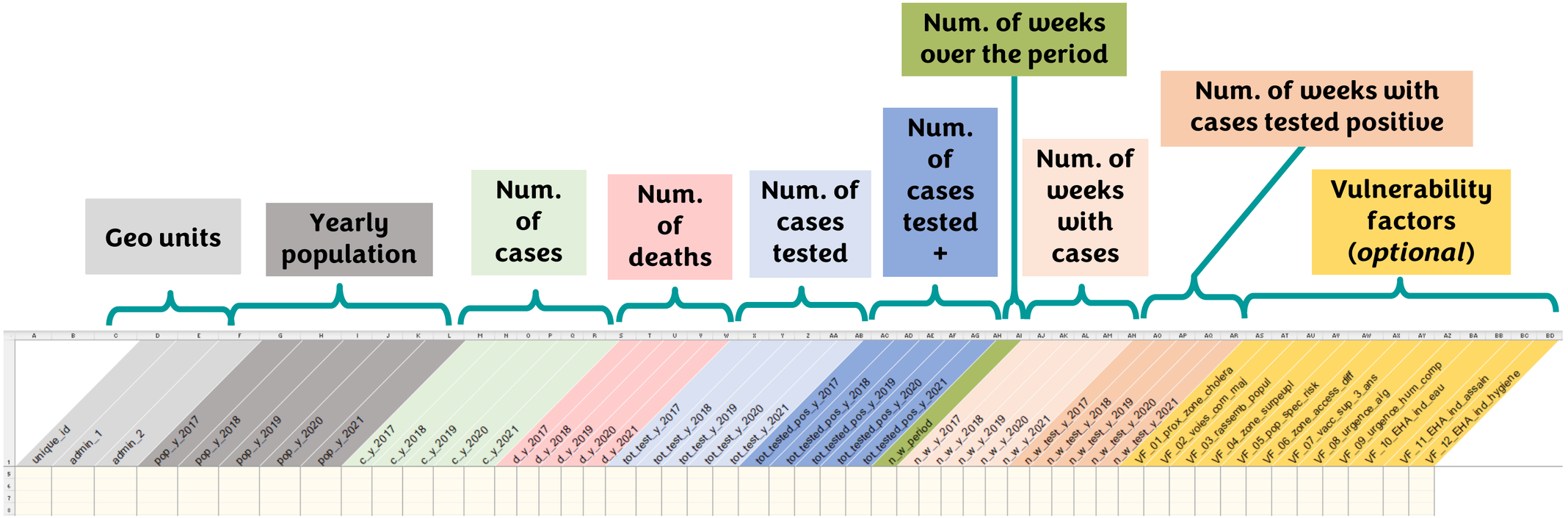
For the PAMI Excel tool to work, the dataset must be formatted in accordance with the PAMI Data Model Template

- Get the Excel file to be used as the PAMI data model

https://tinyurl.com/PAMIcontrol

Structure of the data model

- Each **row** is a geo unit
- Each **column is a variable** for PAMI identification



Do not modify the headings of columns

Except for the year numbers to match your analysis period

Customization of analysis period

By default, the PAMI data model template is for 5 years

If your PAMI analysis period = 5 years

- **Edit the year numbers in the heading of columns**

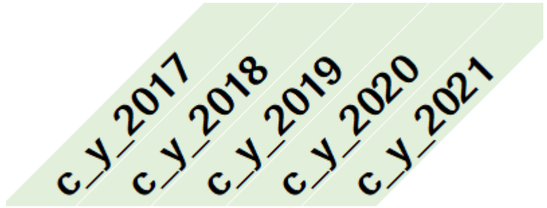
If your PAMI analysis period > 5 years

- **Add years by adding columns**:
  - Add as many columns as needed to match the duration of your analysis period
  - **Follow the template strictly to label each column**

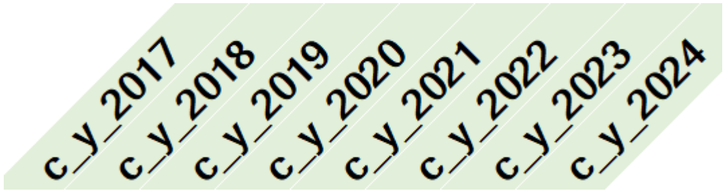
Illustration

Variable number of cases

c_y_YYYY



GTFCC template, 5 years



Customized template, 8 years

Wrap up

To prepare the data:

- Determine the **analysis period** and the **geographic level of PAMIs**
- Compile **retrospective surveillance data** (epidemiological data and testing data) for each geographic unit and for each year of the analysis period
- **Clean the data** and address **missing data**
- **Format the data with the PAMI data model template** so that calculations can be automated in the PAMI Excel tool


Quiz
time

Question 1

- **What could be a potential issue if geo 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

Question 1 – Answer

- **What could be a potential issue if geo 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

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

Question 2 – Answer

- **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

Question 3

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 geo unit

Question 3 – Answer

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 geo unit


Together we can
#Endcholera
GLOBAL TASK FORCE ON
CHOLERA CONTROL