PAMIs for cholera elimination

Module 2

GLOBAL TASK FORCE ON
CHOLERA CONTROL

PAMIs for cholera elimination

Data preparation

What will you learn?

- How to determine the **geographic level of PAMIs**
- Which **data to compile** to identify PAMIs for elimination
- How to **handle missing data**
- How to use the **data model template**

Data compilation

Photo adapted from: WHO / Fabeha Monir

Data compilation

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

- Data to be compiled include:
  - **Geographic** data
  - **Population** data
  - Retrospective **surveillance** data
  - Data on the presence/absence of **vulnerability factors**

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

Geographic level for data compilation

Prior to compiling any data to identify PAMIs

- Define the **geographic level of PAMIs**
- This is the **geographic level at which all data will be compiled**

The geographic level of PAMIs is country-specific
It is determined by country stakeholders in agreement across multiple sectors

Geographic level of PAMIs

- The following is considered to determine the geographic level of PAMIs:
  - **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")**

Time period for data compilation

Retrospective surveillance data

- **Compiled for at least the past 5 years**
  - Occurrence of **confirmed cholera outbreak(s)**
  - Occurrence of **imported cholera case(s)**

Data on vulnerability factors

- **Compiled for the most recent year**
  - **Most up-to data source** available for each factor
  - Might correspond to a different time period for different vulnerability factors

Selection of vulnerability factors

Relevant vulnerability factors are identified in
consultation between stakeholders from multiple sectors

GTFCC indicative list of generic vulnerability factors

- Is **any factor ON this list not relevant** in the local context?
- Is **any additional factor NOT ON this list relevant** in the local context?

For a refresher on the GTFCC indicative list go to Module 1

Additional vulnerability factors

- Associated with a risk of introduction of cholera, onset or spread of a cholera outbreak
- Expert consultations or scoping reviews

Measurable vulnerability indicators

A measurable vulnerability indicator associated with a data source is defined for each vulnerability factor to assess its presence/absence

Vulnerability factor → **Measurable indicator** → **Data source**

- **Interpretation** of the vulnerability factor in a **reproductible** manner
- **Assessement** of presence/absence of the vulnerability factor in an **objective** manner

- Data driven **measurement** of the presence/absence of the vulnerability factor

Selecting data sources

The best data source is the best compromise for data that are:

- **Recent**
- **Reliable**
- **Available at the required geographic level**
- **Comprehensive**

- The **definition of a measurable indicator** may have to be tweaked (e.g., categories, thresholds)

- Data may only be available at the **upper geographic level**
  - For example, geo units are at admin 3 level but data available at admin 2 level
  - If so, geo units “inherit” the value of the upper level

Measurable vulnerability indicators

Examples

Vulnerability factor	Measurable indicator	Data source
• **High population density**	• Number of inhabitants >1,000 per km²	• Ministry of Demography (2024)
• **Unimproved water**	• > 30% of the population using unimproved service level or surface water or • > 15% of the population using surface water	• Joint Monitoring Programme (JMP) - WHO/UNICEF (2022)

Assessment of vulnerability

The presence/absence of each vulnerability factor in each geo unit is assessed using the selected data source and selected measurable indicator

- In the dataset, **absence of a vulnerability factor** versus **absence of data** to assess the presence/absence of the factor must be differentiated
  - "**Missing**" is for when there is **no data** to assess the presence/absence of the factor
  - "**No**" is for when the **factor is absent**

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** or outliers

Missing data

Addressing missing data is essential to limit bias

- **All missing data should be filled before making any decision on PAMIs**
- Different strategies apply depending on the extent of missing data
  - **Partial**
  - **Substantial**

How to address missing data?

Substantial data missingness

- Data to assess the presence/absence of a vulnerability factor **missing for most geo units**
  - Identify an **alternative data source**
  - **Revise the definition** of the measurable indicator

Partial data missingness

- Data to assess the presence/absence of a vulnerability factor **missing for a few geo units**
  - Conduct an ***ad hoc* survey** to collect missing data
  - Ask **subject matter experts** to provide a qualitative assessment
  - If information remains missing, the presence/absence of the vulnerability factor(s) is assessed qualitatively at the **stakeholder validation**

Data formatting

Photo adapted from: WHO / Christopher Black

PAMI data model template

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

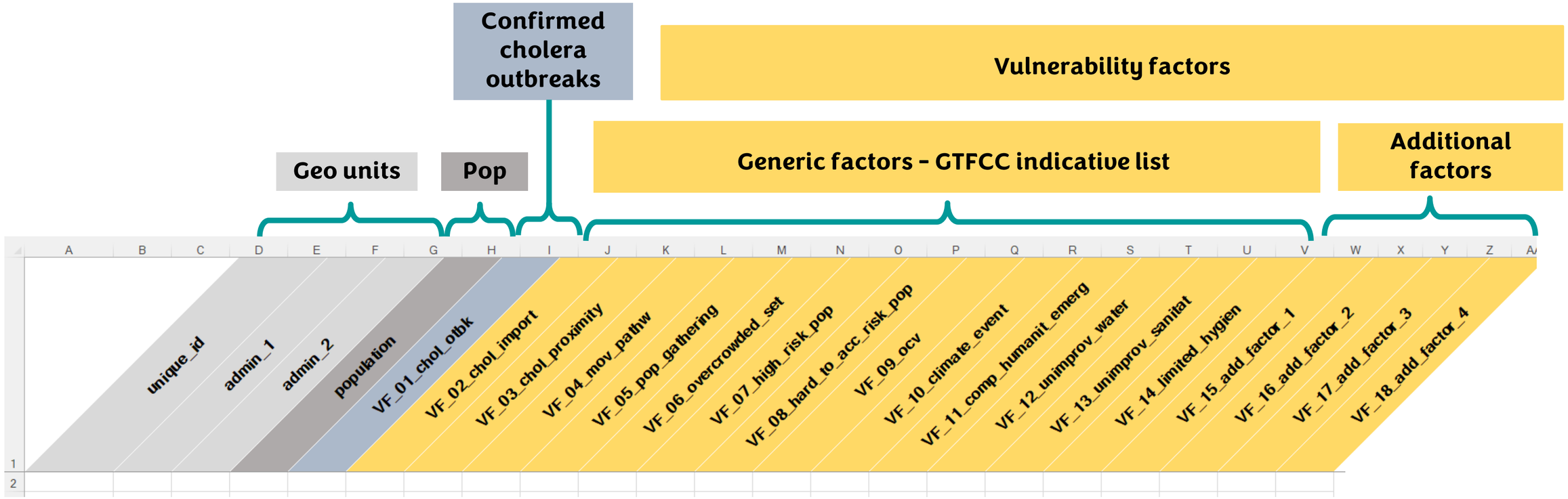
For the PAMI Excel tool to work, the dataset must be formatted in accordance with the PAMI Data Model Template

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

https://tinyurl.com/PAMIelimination

Structure of the data model

- Each **row** is a geo unit
- Each **column** is a variable to identify PAMIs for elimination



Filling the data model

Do not modify the variable names (headings of columns, first line)

- **NO change should be made in variables names**

- If **additional vulnerability factor(s)** is/are included in the analysis:
  - Use the **variables labelled VF_15 to VF_18** without changing their name

Wrap up

To prepare the data

- Determine the **geographic level of PAMIs**
- Compile **retrospective surveillance data** and data on **presence/absence of vulnerability factors**
- **Clean** the dataset and **address missing data**
- **Format** the dataset in accordance 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 data may be too cumbersome to expedite the data collection process

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

d) Significant progress towards achieving cholera elimination may be too slow

Question 2

Why is it essential to fill all missing data?

a) To ensure all geo 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

Question 2 – Answer

Why is it essential to fill all missing data?

a) To ensure all geo 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

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 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 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 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 geo unit


Together we can
#Endcholera
GLOBAL TASK FORCE ON
CHOLERA CONTROL