Optimizing Schistosomiasis MDA Implementation

Sub-Implementation Data Analysis Tool

User Guide

This document aims at providing users a guide on how to use the "ESPEN sub-district data analysis tool for optimizing schistosomiasis MDA implementation Version 1 Sept 2019". Based on joint application packages (JAP) received from countries, preventive chemotherapy for schistosomiasis using praziquantel distribution is usually implemented at the district level. This is over a larger spatial scale than may be needed, as schistosomiasis transmission is typically localized to water contact site catchment areas. There are growing concerns around the efficiency of distribution of donated praziquantel, and sub-optimal utilization of site-level data. Therefore, review of sub-district level data may help to better determine appropriate implementation units. In order to help countries to review their existing data, a data analysis tool has been designed. Twenty-four countries have been trained during two workshops. This document aims at providing users a guide on how to use the tool. Any feedback on this document may be communicated to mwinzip@who.int.

World Health Organization
REGIONAL OFFICE FOR Africa

ESPEN EXPANDED SPECIAL PROJECT FOR ELIMINATION OF NEGLECTED TROPICAL DISEASES

Version 1.0 - 20190910

September 2019

Contents

Acknowledgements

ESPEN would like to express special thanks to all the following individuals or institutions who contributed to the tool and the user guide:

ESPEN CONTRIBUTORS

Pauline Mwinzi
Honorat Zoure
Lamine Diawara
Maria Rebollo Polo

LSHTM CONTRIBUTORS

Rachel Pullam
Jorge Cano
Hope Simpson
David Kennedy

CONSULTANTS

Eugene Ruberanziza
Boniface Ekoue Kinvi
Dieu Merci Dwe
Jorel Boulinzann
Djima Adam Mama
Ange Elvis Aba
Mamadou Serme
Benjamin Tinkitina
Lum Brunhilda
Joyline Marimo
Robson Tirivangani Mukwiza
Levison Nkhoma
Trevor Muchabaiwa

PARTICIPATING ESPEN PARTNERS

GSA
SCI
FHI 360

Abbreviations

BIU	Blood in urine questionnaire test
CCA	Circulating Cathodic Antigen test for S. mansoni detection
ESPEN	Expanded Special Project for Elimination of NTDs
IU	Implementation Unit
JAP	Joint Application Package
JRSM	Joint Request for Selected PC Medicines
MDA	Mass Drug Administration
PC	Preventive Chemotherapy
PZQ	Praziquantel
SAC	School age children
SN	Serial Number
WHO	World Health Organization

Working Definitions

Sub-district level or Sub-IU level: as used in this analysis, sub-district refers to the lowest administrative level possible below the current implementation level/unit (IU), for which shape files, epidemiological and demographic data are available for analysis. *Sub-district, sub-IU and sub-unit are used interchangeably in the document.*

List of tables

List of figures

1 Introduction

Based on joint application packages (JAP) received from countries, preventive chemotherapy for schistosomiasis using praziquantel distribution is usually implemented at the district level. This is over a larger spatial scale than may be needed, as schistosomiasis transmission is typically localized to water contact site catchment areas. There are growing concerns around the efficiency of distribution of donated praziquantel, and sub-optimal utilization of site-level data. Therefore, review of sub-district level data may help to better determine appropriate implementation units. In order to help countries to review their existing data, a data analysis tool has been designed, *"ESPEN sub-district data analysis tool for optimizing schistosomiasis MDA implementation Version 1 Sept 2019".* The data analysis tool is designed in an Excel workbook with automated tasks to process the data, view the results and make decisions. Twenty-four countries have been trained during two workshops in July and August 2019. This document aims at providing users with a guide on how to use the tool.

This guide is structured in three chapters. The first chapter describes the data requirements and the overall structure of the workbook. The second chapter describes the data processing tasks and the third chapter presents the output summary reports and the final output.

2 Data requirement and structures

The workbook contains 5 types of worksheets:

- the data dictionary worksheet,
- the instruction worksheet,
- the input worksheets (demo_data, epi_data, geo_data, Endemicite_JRSM),
- the analysis and output worksheets (Sub_District_Data_Analysis, Sub_District_Summary),
- the output summary worksheets (1_cat_end, 2_changes, 3_adequacy, 4_Pop_req_d_sd, 5_PZQ_req_d_sd, 6_Advantages, 7_Pop_Req_PC)

2.1 Data dictionary

The data dictionary describes the variables used in the workbook. The description includes their titles (group, sub-group, column number, name and full description) and the input mode (inputted or calculated).

2.2 Instructions worksheet

The instruction worksheet contains quick instruction on the procedures. It also the place where the user will provide some key information such as the total number of sub-units.

2.3 Input worksheets

There are four input worksheets that constitute the data sources for the inputted data. They are used at different stages of the calculations.

2.3.1 Demographic data

The worksheet demo_data is the data source for the demographic data. It must contain the complete list of all sub-IUs in the country whether endemic or not for schistosomiasis. The data items are as in the table below.

Table 1: Demographic data structure

Category	Column number	Variable	Description	Input Mode
Geography	1	SN	Incremental number of the sub-IU	Inputted
	2	Country	Name of the country	Inputted
	3	ISO2	ISO 2 characters code of the country	Inputted
	4	Admin_1	Administrative level 1 (usually the level immediately above the IU)	Inputted
	5	Admin_2 (IU)	Administrative level representing the IU (implementation unit)	Inputted
	6	ESPEN_IU_ID	ESPEN internal ID for the IU	Inputted
	7	Admin_3 (sub-IU, sub-district)	Sub-IU name (may not be always the admin 3, but much lower – must be level below the IU chosen for the sub-IU implementation)	Inputted
Demography	8	Year of population	Year of population	Inputted
	9	Total Population	Total Population	Inputted
	10	SAC Population	School age children Population	Inputted
	11	Adult Population	Adult Population	Inputted
	12	Percentage of SAC (%)	Percentage of school age children (%)	Inputted
	13	Percentage of adults (%)	Percentage of adults (%)	Inputted

Notes

The demographic dataset is the first data source and as such must be complete as much as possible. The SN in the sequential number generated for each sub-IU. It is recommended to sort alphabetically the dataset by Admin_1, Admin_2 and Admin_3 before generating the SN.

2.3.2 Epidemiological data

The worksheet epi_data is the data source for the epidemiological data. It must contain all available site level survey data. The data items are as in the table below.

Table 2: Epidemiological data structure

Category	Column number	Variable	Description	Input Mode
Geography	1	SN	Incremental number of the data row	Inputted
	2	Country	Name of the country	Inputted
	3	ISO2	ISO 2 characters code of the country	Inputted
	4	Admin_1	Administrative level 1 (usually the level immediately above the IU)	Inputted
	5	Admin_2 (IU)	Administrative level representing the IU (implementation unit)	Inputted
	6	ESPEN_IU_ID	ESPEN internal ID for the IU	Inputted
	7	Admin_3 (sub-IU, sub-district)	Sub-IU name (may not be always the admin 3, but much lower – must be level below the IU chosen for the sub-IU implementation)	Inputted
	8	Site	Name of the survey site	Inputted
	10	Latitude	Latitude in decimal degrees	Inputted
	11	Longitude	Longitude in decimal degrees	Inputted
Epidemiology	9	Survey Type	The type of survey	Inputted
	12	Examined	Number of people examined	Inputted
	13	Positive	Number of people positive	Inputted
	14	Prevalence (%)	Prevalence rate calculated as percentage (positive/examined*100)	Inputted
	15	Methods	Methods of diagnostic	Inputted
	16	Year of survey	Year of survey	Inputted
	17	Species	Species diagnosed	Inputted
	18	Diagnostic Type	Type of diagnostic methods (parasitological, serological, qualitative, etc.)	Inputted

Notes

The column number of each variable is used in the calculation. The diagnostic methods been considered in the calculation are 8 grouped in 3 types and must be written exactly as in the table below.

Table 3: Coded name of the diagnostic methods and types

Diagnostic Type	Diagnostic Methods	Species
Parasitology	Urine Filtration	Urogenital
	Urine Sedimentation	Urogenital
	Kato-Katz	Intestinal
	Other microscopy	Intestinal
Clinical	Dipstick	Urogenital
	BIU	Urogenital
	Urine Other	Urogenital
Serology	CCA	Intestinal

2.3.3 Endemicity data reported in the latest JRSM

The worksheet Endemicity_JRSM contains the endemicity category reported in the most recent JRSM for the IU. The data items are as in the table below.

Table 4: Structure of the endemicity data reported by the latest JRSM

Category	Column number	Variable	Description	Input Mode
Geography	1	SN	Incremental number of the sub-district	Inputted
	2	Country	Name of the country	Inputted
	3	ISO2	ISO 2 characters code of the country	Inputted
	4	Admin_1	Administrative level 1 (usually the level immediately above the IU)	Inputted
	5	Admin_2 (IU)	Administrative level representing the IU (implementation unit)	Inputted
	6	ESPEN_IU_ID	ESPEN internal ID for the IU	Inputted
	7	District ID	Internal IU ID used globally in CDS_NTD SQL Server database	Inputted
Epidemiology	8	Year	Reporting Year	Inputted
	9	Endemicity	Endemicity full description	Inputted
	10	Short description	Endemicity short description (Not endemic, Low, Moderate, High, Unknown)	Inputted

Notes

There are 5 categories of endemicity descriptions that should be coded in the column of short description as Not endemic, Low, Moderate, High and Unknown. The description used in the calculation is the short description.

2.3.4 The neighboring sub-districts

The worksheet Geo_Data contains the proximity relationship between the sub-IUs. Each row in the worksheet indicate a link between 2 adjacent (sharing of border). Each sub-IU will appear the number of times it has adjacent sub-IUs. The table below shows the structure of the dataset

Table 5: Structure of the neighboring sub-units

Category	Column number	Variable	Description	Input Mode
Country	1	SN	Incremental number of the data row	Inputted
	2	country	Name of the country	Inputted
	3	ISO2	ISO 2 characters code of the country	Inputted
Sub-IU Information	4	ADM1	Administrative level 1 (usually the level immediately above the IU) of the sub-IU	Inputted
	5	ADM2	Administrative level representing the IU (implementation unit) of the sub-IU	Inputted
	6	ADM3 (sub-IU)	Sub-IU name	Inputted
Neighboring sub-IU Information	7	ADM3_neighbourhood (sub-IU)	Adjacent sub-IU name	Inputted
	8	ADM2_neighbourhoog	Administrative level representing the IU (implementation unit)	Inputted
	9	ADM1_neighbourhood	Administrative level 1 (usually the level immediately above the IU) of the adjacent sub-IU	Inputted

2.4 Analysis and Output Worksheets

There are 2 analysis and output worksheets. The analysis is done in 3 steps. The first 2 steps are achieved in the first analysis worksheet named Sub_District_Data_Analysis and the third step is performed in the second analysis worksheet named Sub_District_Summary.

2.4.1 District level and sub-district level Analysis

The District and sub-district levels analysis is performed in the worksheet named "Sub_District_Data_Analysis". Its structure is the most complex and divided into different sections: geography, demography, district level implementation parameters, sub-district level implementation parameters, comparative advantages between district and sub-district implementation.

2.4.1.1 Geography section

The geography section is the same as in the demographic and epidemiological worksheets. It is the replication of the same information in the demographic data worksheet.

2.4.1.2 Demography section

The demographic data section replicates the structure of the demographic data in the demographic data worksheet. The data filled in that section are referenced to the same data in the demographic worksheet.

2.4.1.3 Epidemiology section

The epidemiological data values are calculated from the data stored in the epidemiological data worksheet. From the site level data, global prevalence is calculated either for the district or the sub-district for each diagnostic method for which data is available. The global prevalence is then translated into endemicity category.

The epidemiological section is divided into groups: one for the district level data and another for the sub-district level data. The 2 groups have the same variables.

2.4.1.3.1 Epidemiological Data and Parameters for District Level Implementation

The table below shows the structure of variables calculated for the district level implementation.

Table 6: Structure of the epidemiological data on the district

Category		Column number	Variable	Description	Input Mode
Inclusion Exclusion		12	Included in comparisons	Indicate the sub-IU that have demographic and epidemiological data to be included in the comparisons	Calculated
Epidemiological Data and Parameters for District Level Implementation	**Parasitological tests**	13	Prev_sh_UF (IU)	Prevalence by urine filtration of the district	Calculated
		14	Prev_sh_US (IU)	Prevalence by urine sedimentation of the district	Calculated
		15	Prev_KK (IU)	Prevalence by Kato-Katz of the district	Calculated
		16	Prev_oth_micr (IU)	Prevalence by other microscopy (Intestinal) of the district	Calculated
	Clinical and qualitative tests	17	Prev_dip (IU)	Prevalence by dipstick of the district	Calculated
		18	Prev_BIU (IU)	Prevalence by BIU questionnaire of the district	Calculated
		19	Prev_urine_oth (IU)	Prevalence by other urine test of the district	Calculated
	Serological tests	20	Prev_sm_CCA (IU)	Prevalence by CCA of the district	Calculated
	District endemicity parameters	21	Methods (IU)	Diagnostic methods of final prevalence of the district	Calculated
		22	Prev_sch (IU)	Final prevalence of the district	Calculated
		23	Number of sites (IU)	Number of sites for the final prevalence of the district	Calculated
		24	Year (IU)	Year of the final prevalence of the district	Calculated
		25	Endemicity category calculated (IU)	Endemicity category by the final prevalence of the district	Calculated
		26	Endemicity category in district (by latest JRSM)	Endemicity category from the most recent JRSM	Calculated
	Estimations of population and drugs	27	SAC estimates (IU endemicity)	Estimations of SAC by the prevalence of the district	Calculated
		28	Adults estimates (IU endemicity)	Estimations of adults by the prevalence of the district	Calculated
		29	Drugs estimates for SAC (IU endemicity)	Estimations of PZQ for SAC by the prevalence of the district	Calculated
		30	Drugs estimates for adults (IU endemicity)	Estimations of PZQ for adults by the prevalence of the district	Calculated

Notes

The column 12 can be “Yes” or “No”. The “Yes” value indicates which sub-IUs will be included in the comparison analysis. It is “Yes” under the following conditions:

1. The sub-district has population data
2. The sub-district has epidemiological data that led to its endemicity classification
3. The sub-district epidemiological data has the same parameters (diagnostic methods and year of survey) as the district epidemiological data

The columns 13 to 20 are the calculated prevalence for the different diagnostic methods and the column 21 to 26 are the parameters of the endemicity category calculated. The column 26 is filled with the latest endemicity reported in JRSM.

The columns 27 to 30 are filled with the estimations of population and praziquantel based on the endemicity category.

2.4.1.3.2 Epidemiological Data and Parameters for Sub-district Level Implementation

The table below shows the structure of variables calculated for the sub-district level implementation.

Table 7: Structure of the epidemiological data on the sub-district

Category		Column number	Variable	Description	Input Mode
Epidemiological Data and Parameters for District Level Implementation	**Parasitological tests**	31	Prev_sh_UF (sub-IU)	Prevalence by urine filtration of the sub-district	Calculated
		32	Prev_sh_US (sub-IU)	Prevalence by urine sedimentation of the sub-district	Calculated
		33	Prev_KK (sub-IU)	Prevalence by Kato-Katz of the sub-district	Calculated
		34	Prev_oth_micr (sub-IU)	Prevalence by other microscopy (Intestinal) of the sub-district	Calculated
	Clinical and qualitative tests	35	Prev_dip (sub-IU)	Prevalence by dipstick of the sub-district	Calculated
		36	Prev_BIU (sub-IU)	Prevalence by BIU questionnaire of the sub-district	Calculated
		37	Prev_urine_oth (sub-IU)	Prevalence by other urine test of the sub-district	Calculated
	Serological tests	38	Prev_sm_CCA (sub-IU)	Prevalence by CCA of the sub-district	Calculated
	District endemicity parameters	39	Methods (sub-IU)	Diagnostic methods of final prevalence of the sub-district	Calculated
		40	Prev_sch (sub-IU)	Final prevalence of the sub-district	Calculated
		41	Number of sites (sub-IU)	Number of sites for the final prevalence of the sub-district	Calculated
		42	Year (sub-IU)	Year of the final prevalence of the sub-district	Calculated
		43	Endemicity category calculated (sub-IU)	Endemicity category by the final prevalence of the sub-district	Calculated
	Estimations of population and drugs	44	SAC estimates (sub-IU endemicity)	Estimations of SAC by the prevalence of the sub-district	Calculated
		45	Adults estimates (sub-IU endemicity)	Estimations of adults by the prevalence of the sub-district	Calculated
		46	Drugs estimates for SAC (sub-IU endemicity)	Estimations of PZQ for SAC by the prevalence of the sub-district	Calculated
		47	Drugs estimates for adults (sub-IU endemicity)	Estimations of PZQ for adults by the prevalence of the sub-district	Calculated

The notes above on the district level are the same for the sub-district level implementation.

2.4.1.4 Adequacy and inadequacy of district level implementation

The last part of the analysis worksheet is filled with information that measure the inadequacy of the district level implementation. It indicates whether a treatment strategy under district level implementation compared to sub-district level is adequate or not (under treating or over treating). It also indicates the gaps in under treatment as well as the wastage in over treatment.

Table 8: Structure of the calculated data on the treatment strategy adequacy

Category	Column number	Variable	Description	Input Mode
adequation of the district level implementation	48	Treatment Strategy Adequacy	Number Treatment adequacy	Calculated
	49	SAC adequately treated	Population of SAC treated adequately	Calculated
	50	SAC missing treatment	Population of SAC missing treatment	Calculated
	51	SAC unnecessary treated	Population of SAC unnecessary treated	Calculated
	52	Unclaimed drugs	Unclaimed PZQ	Calculated
	53	Drugs misused	Mis-used PZQ	Calculated

2.4.2 Sub-district Summary worksheet

The sub-district summary worksheet named “Sub_District_Summary” is a compacted dataset from the global analysis. It has additional data processing capability. Its structure has 5 sections: geography, demography, epidemiology and estimations.

2.4.2.1 Geography section

The geography section is the same as in the other worksheets that have the same section. The data filled in this section are referenced to the same data in the demographic worksheet.

2.4.2.2 Demography section

The demographic data section replicates the structure of the demographic data in the demographic data worksheet. The data filled in that section are referenced to the same data in the demographic worksheet.

Table 9: Structure of the summary sub-district data

Category	Column number	Variable	Description	Input Mode
Geography	1	SN	Incremental number of the sub-unit	Inputted
	2	Country	Name of the country	Inputted
	3	ISO2	ISO code 2 characters of the country	Inputted
	4	Admin_1	Administrative level 1 (usually the level immediately above the IU)	Inputted
	5	Admin_2 (IU)	Administrative level representing the IU (implementation unit)	Inputted
	6	ESPEN_IU_ID	ESPEN internal identifier for the IU	Inputted
	7	Admin_3 (sub-IU, sub-district)	Subunit name (may not always be level 3, but much lower - must be lower than IU)	Inputted
Demography	8	Year of population	Year of population	Inputted
	9	Total Population	Total population	Inputted
	10	SAC Population	School-age children population	Inputted
	11	Adult Population	Adult population	Inputted
Epidemiology	12	Endemicity category calculated by the IU data	Category of endemicity calculated with IU data	Calculated

Category	Column number	Variable	Description	Input Mode
	13	Endemicity category reported by the latest JRSM	Category of endemicity reported in the last JRSM	Inputted
	14	Endemicity category calculated by the sub-IU data	Category of endemicity calculated with sub-IU data	Calculated
	15	Final Decision	Final decision to determine the endemicity of the sub-IU	Inputted
	16	Final Endemicity Category	Category of final endemicity	Inputted/Calculated
Estimations	17	Estimations of SAC	EAS estimates	Calculated
	18	Drugs estimations for SAC	PZQ estimates for EAS	Calculated
	19	Estimations of adults	Adult estimates	Calculated
	20	Drugs estimations for adults	PZQ estimates for adults	Calculated
Treatment Plan	21	2019	2019	Inputted
	22	2020	2020	Inputted
	23	2021	2021	Inputted

2.4.2.3 Epidemiology section

The epidemiology section has 2 groups: the first group has 3 columns (Category of endemicity calculated with IU data, Category of endemicity reported in the last JRSM, Category of endemicity calculated with sub-IU data) that report endemicity calculated in the global analysis worksheet. The second group has 2 columns:

- Final decision to determine the endemicity of the sub-IU: This column invite the user to input the decision criteria to be used to determine the final endemicity of the sub-district. It has a dropdown list that offers 6 options.

- Category of final endemicity: The final endemicity calculated based on the final decision is filled in this column.

2.4.2.4 Estimation section

This section calculates the school age children and the adult population that need treatment according to the final endemicity.

3 Data Processing

3.1 Data Preparation

The data preparation task consists of filling in the input data worksheets with the available datasets for the analysis. Four worksheets are concerned by this task: demographic data worksheet, epidemiological data worksheet, JRSM data worksheet, and the neighboring subunits worksheet. The first two are mandatory while the third and the fourth ones are optional (they are only optional if not available and mandatory when available). Each dataset if available must be prepared and formatted strictly in conformity with the worksheet structures as described in the first chapter of this guide.

3.1.1 Filling in the input worksheets

There are no standard methodologies provided in this guide about how to fill in the worksheets. You can transfer your dataset from its current storage to your data analysis template and work on it or you can work on it first from its current storage before transferring it to the analysis template. Whichever way you choose, below some recommendations:

- Take enough time to prepare the dataset in a separate worksheet or in any other application where the data might have been previously stored (e.g. MS access database, SPSS, Stata, CSPro, DHIS2, etc.) or where you fill more comfortable, especially if the data is not yet analyzed as required by the analysis and needs some reworks before (e.g. calculate site prevalence from individual records, linking demographic or epidemiological data to shapefiles in order to link them to their subunits)
- Format the dataset in the same structure (column ordering) as in the destination worksheet
- No full blank row is allowed
- No full blank column is allowed except if it is in the structure, but the data is not available
- No merged cell is allowed in the dataset
- Cell formatting such as background and fore color, font properties, framing are allowed.
- Avoid formatting the worksheet out of the actual data range
- You can keep the column titles (description) as they are, or you can rewrite them in your own way. But you must keep in mind that the data type in the column should not change (e.g. In the demographic data worksheet, the name of the column 7 is “Admin_3 (sub-IU, sub-district)” and meant to store the sub-IUs names. You can change that name title to sub-district, Sub-county, ward, parish etc. if that is the name of the sub-IUs in your country and knowing that you will fill in the column the subunits chosen as the sub implementation unit).

3.1.2 Data cleaning and quality check

Once, you have filled in the worksheet either in the data analysis template or in another workbook or tool, you can proceed to the next important step which is the data cleaning. The data cleaning will mostly concern the data values and characteristics.

3.1.2.1 Geographical unit names and spellings

- Make sure that all geographical unit (admin1, admin2, admin3) names are written correctly and in the same way wherever they appear. They appear all in the fourth input worksheets
- Remove any useless spaces at the beginning or at the end of the names (it happens very often that spaces are left inadvertently at the extremities of words)

- Compound names should be separated by a straight dash (-) and not an underscore (_). No space is needed on left and right of the dash.
- Some use a slash (/) to separate compound names and this can conflict with some data processing tools where slashes are used as a special character. So, avoid the use of the slash as much as possible in compound names.

- If you have numbered names such as district 1, district 2 and so one, it is recommended to use the suffix 1, 2, 3, 4, etc. rather than the roman numerals I, II, III, IV etc.

- When different units have an identical name (this happen very often in big countries and at the lower level units), make sure that you rename those units uniquely by (1) adding a different suffix to the name, or (2) precede the name with the parent unit name separated by a dash.

3.1.2.2 Numerical values

Numerical values must contain only digits. No alphabetical character in the value. If you want to separate the thousands, please use the built-in number formatting tools.

3.1.2.3 Text values

Text values are often data characteristic and rarely data values as such. In all cases some precautions are to be taken.

- Most of the above rules for the geographical unit names apply to other data characteristics
- Same characteristic names may be referred to the same state, so there is no need to differentiate them. Rather they must be written the same (e.g. Urine filtration for diagnostic method should not be written differently)

3.1.3 Data coding

A bit of data codding is required for the completeness of the data processing

3.1.3.1 Sorting

It is recommended that you sort the dataset in alphabetical order (A to Z) by the geographical unit (admin 1, admin 2, admin 3)

3.1.3.2 Serialization

All the input worksheet contains a first column “SN” meant to serialize the rows. After you sort the dataset, you fill in the SN column with an incremental number from 1 to the number of rows in the dataset. However, if you have any serial number or unique identifier for the data row that you would like to keep, you may store it in the SN column.

The SN value is mandatory and should not be left empty for any row.

3.1.3.3 Coded values

Some values must be coded in a given way to be interpreted by the analysis tool. Such data items are:

- Diagnostic methods, type and disease type
  The names of the diagnostic methods and types must be written as presented in the table below. The disease type in the worksheet is the column "Species". The term type of disease may be more appropriate since the qualitative tests do not detect the species.

Table 10: Coded values of diagnostic methods and types

Diagnostic type	Diagnostic method	Disease Type
Parasitological	Urine Filtration	Urogenital
Parasitological	Urine Sedimentation	Urogenital
Parasitological	Kato-Katz	Intestinal
Parasitological	Other microscopy	Intestinal
Clinical and qualitative	BIU	Urogenital
Clinical and qualitative	Dipstick	Urogenital
Clinical and qualitative	Other Urine Test	Intestinal
Serological	CCA	Intestinal

- Endemicities categories
  The endemicity categories values are coded as presented in the table below. This codification is required when preparing the JRSM dataset.

Table 11: Coded values of endemicity categories in the JRSM dataset

Numerical code in JRSM	Description in JRSM	Coded description in the analysis tool
0	Non-endemic	Not endemic
1	Low prevalence (less than 10%)	Low
2	Moderate prevalence (10%-49%)	Moderate
3	High prevalence (50% and above)	High
4	Status unknown	Unknown
5	Endemic, prevalence unknown	Unknown

3.1.4 Dealing with missing or unavailable data

Not all the data columns are used during the calculations. The mandatory columns are described below.

- Geographical data
  The columns representing the 3 levels (admin 1, admin 2 and admin 3) are mandatory in all of the four input datasets. If any of them is missing, in the demographic data, the calculations will not be performed for the entire row, and if any of them is missing in the other datasets, the data in those rows will not be considered in the calculation.

- The demographic data are also important to calculate the estimations of people requiring PC. But they are not mandatory for the prevalence calculation

- In the epidemiological dataset, in addition to the geographical data, the following data columns are mandatory: Examined, Positive, Prevalence, Diagnostic methods and Year. If all of them are missing, the row will definitely be ignored. But if some of them are missing, the

quality of the result will be affected. The consequence will be that a data row which primarily is of a good quality may be downgraded (because of a missing column) in the profit of another data row of a low quality. It is very important to provide all the information required for the data that you mostly trust of a good quality.

3.2 Calculations

Once the input data worksheets are prepared (structured and coded as required), the calculations can be performed. Let recall again that the prevalence calculation will yield results only if epidemiological data are available and estimations are done if populations data is available.

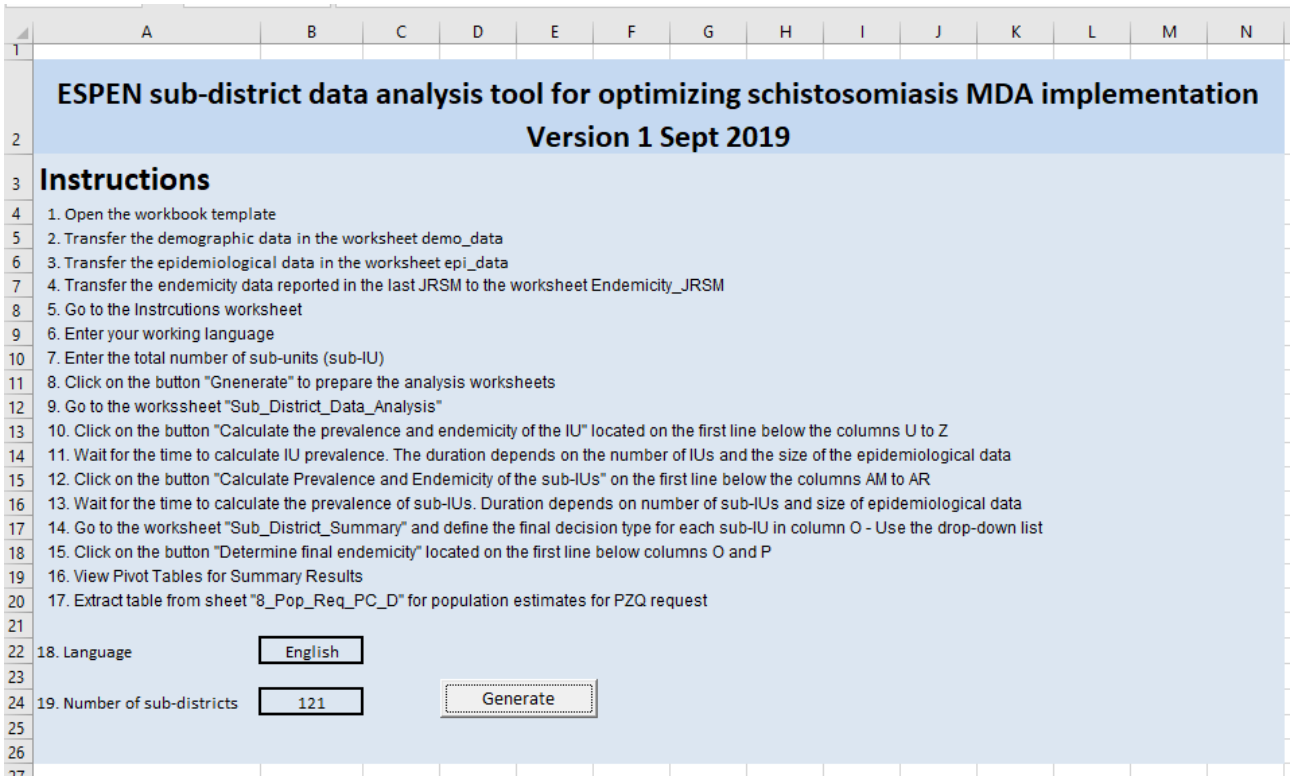
Before running the calculation, you must first generate the two analysis workbooks.

3.2.1 Generate the analysis worksheets

Go to the worksheet Instructions

When you come to this worksheet at this stage, it means you performed step 1 through to 5.

Figure 1: Instructions page



Proceed now to steps 6, 7 and 8 as follow:

- Select your working language
- Enter the total number of sub-units that is in the demographic worksheet
- Click the button Generate to generate the worksheet.

3.2.2 Perform the calculations

There are four calculations processes: two in the Sub_District_Data_Analysis worksheet and two in the Sub_District_Summary worksheet.

3.2.2.1 Calculation in the Sub_District_Data_Analysis worksheet

In this worksheet, you can calculate the prevalence and endemicity category for the sub district using the district epidemiological data and you can also calculate the same using the sub district epidemiological data.

3.2.2.1.1 Prevalence calculation methodology

- Prevalence calculation

Prevalence are calculated for all diagnostic methods for which data are available and met data requirements. The figure below shows the algorithm for prevalence calculation.

- Endemicity categorization

Endemic categories are determined according to WHO guidelines. They are determined for the diagnostic methods included in the guidelines, namely parasitological tests, hematuria, and the serological test CCA. When prevalence is calculated for several diagnostic methods, the choice of prevalence to be considered is made on the basis of the quality grid of diagnostic methods. The Table below contains the diagnostic methods with their degree of reliability.

Figure 2: Prevalence calculation flowchart

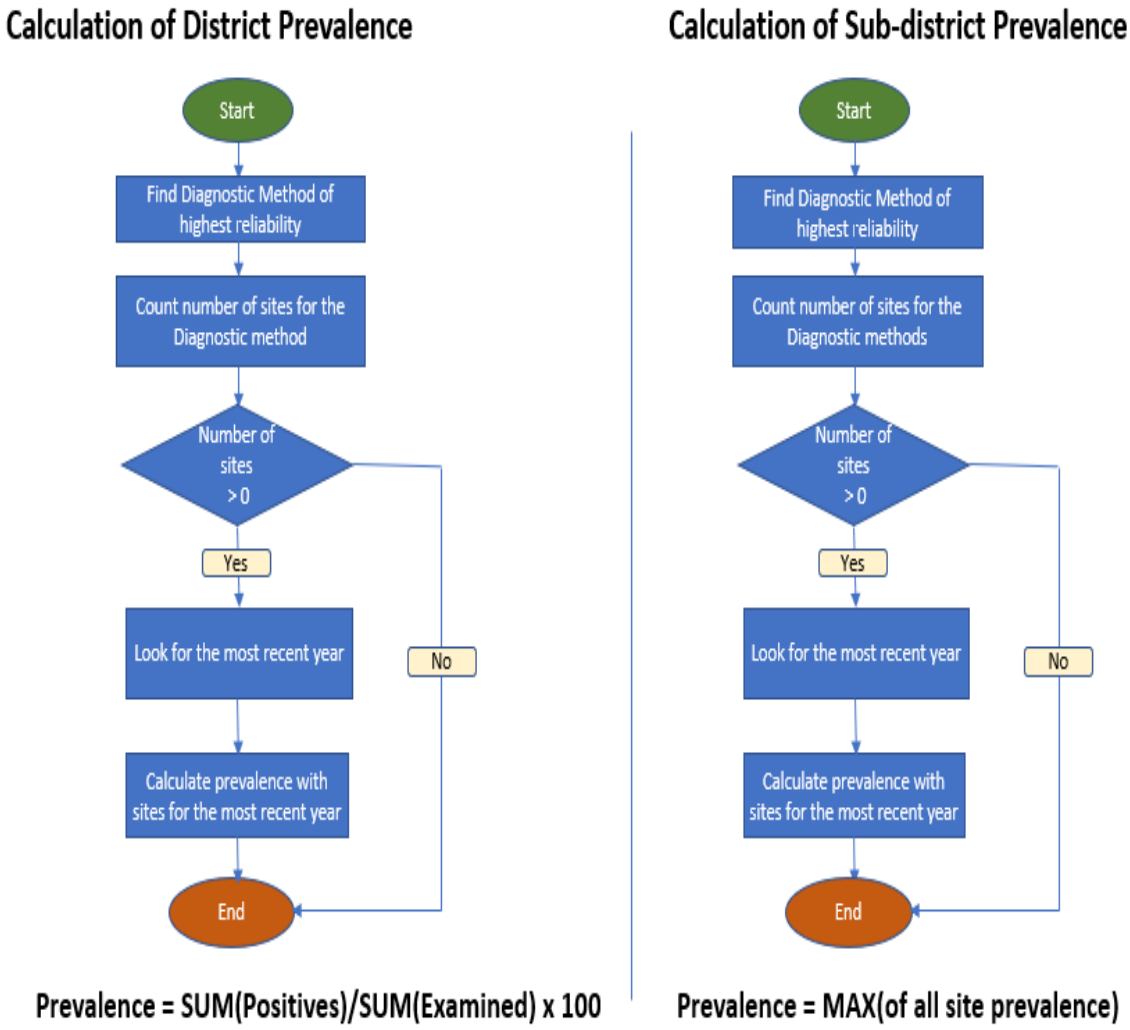


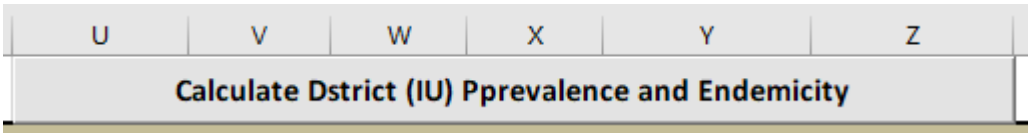
Table 12:Diagnostic Methods by reliability

Diagnostic method	Quality
Urine Filtration	1
Urine Sedimentation	1
Kato-Katz	1
Other microscopy	2
Dipstick	3
BIU	3
Urine Other	4
CCA	4

Rank of diagnostic methods from highest to lowest reliability

3.2.2.1.2 Calculate the IU prevalence and endemicity

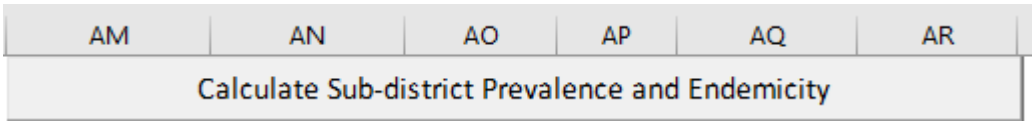
To do the calculation with the district epidemiological data, click the button "Calculate the prevalence and endemicity of the IU" located on the first line below the columns U to Z.



The calculation will start, and the worksheet will scroll down as the calculation progress and you can follow on the screen. Wait for the time it will take. This depends on the number of IUs and the size of the epidemiological data. You can still use your computer for other tasks but avoid any other works in Excel.

3.2.2.1.3 Calculate the sub-district prevalence and endemicity

To do the calculation with the sub-district epidemiological data, click the button "Calculate Prevalence and Endemicity of the Health Area" on the first line below the columns AM to AR



The calculation will start. The worksheet will scroll down as the calculation progress and you can follow on the screen. Wait for the time it will take. This depends on the number of IUs and the size of the epidemiological data. You can still use your computer for other tasks but avoid any other works in Excel.

3.2.2.1.4 Inspecting the calculation results

Many factors will affect the completeness of the results.

- No prevalence calculated for some IUs or sub-IUs
  The reasons may be the followings:
  No epidemiological data available
  Geographical data (unit names) do not match in the demographic and epidemiological data
- Prevalence is calculated but endemicity category is missing

The reason can be that the diagnostic method for which the prevalence is calculated is not recommended in the WHO guidelines

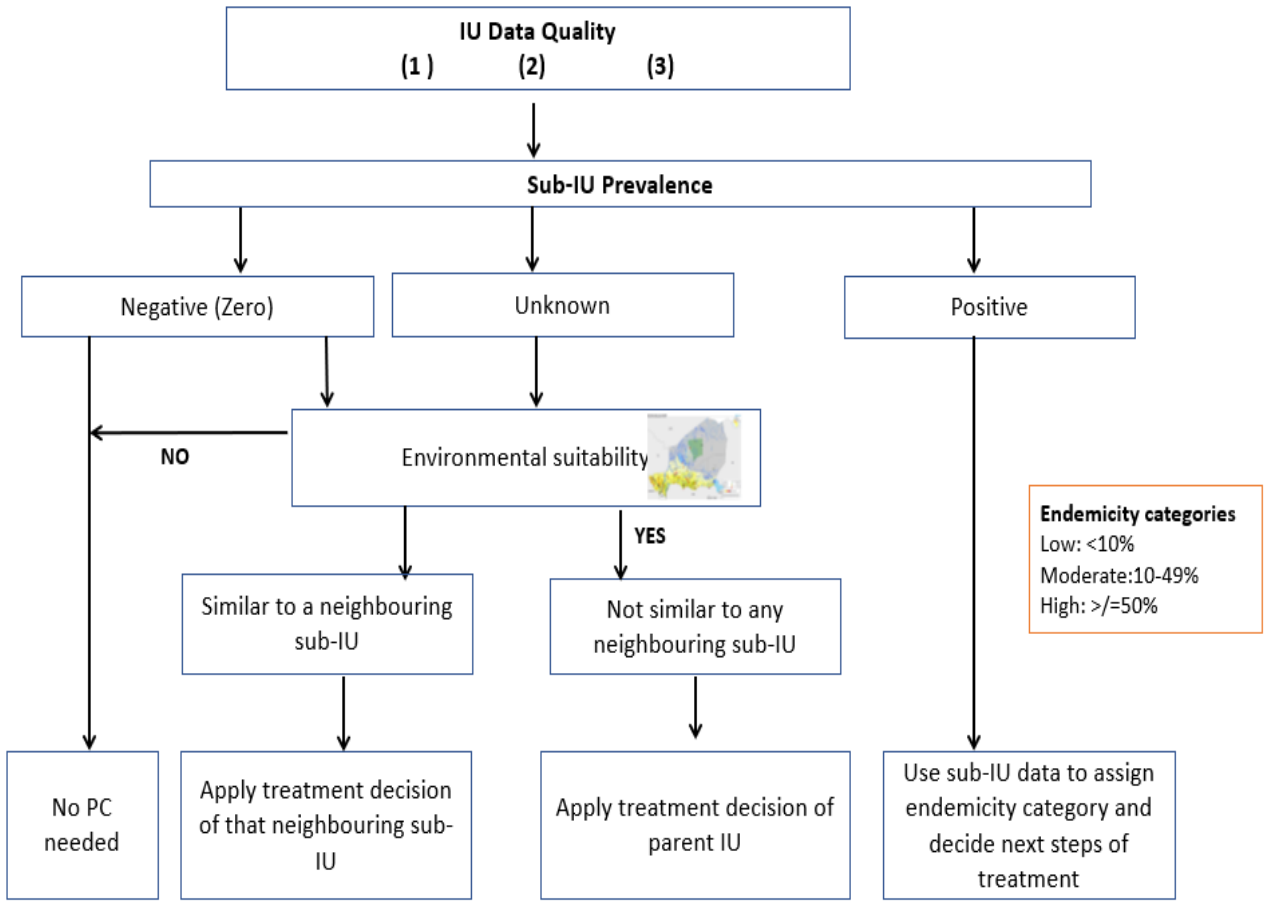
3.2.2.2 Calculation in the Sub_District_Summary worksheet

This worksheet is a summary of the previous one in the sense that only interpretable (endemicity categories) results are replicated. The calculation here consists of determining the final endemicity category for the sub-district in 2 steps first defining the final decision and then determine the endemicity category. Four options are available for the final decision: (1) use the sub-district endemicity calculated with the sub-district epidemiological data, (2) use the district endemicity calculated with the district epidemiological data, (3) use the district endemicity reported in the last JRSM and (4) use the endemicity of a neighboring sub-district. In case none of those 4 options is applicable, there is a last option to differ the decision until further assessment.

3.2.2.2.1 Assign the final decision

The final decision is made in concordance with the flowchart on the figure 2 below.

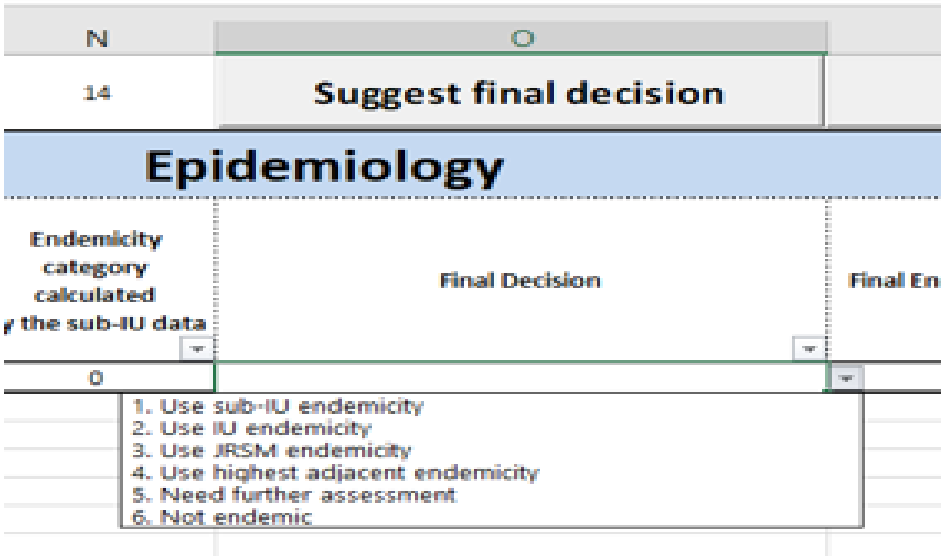
Figure 3: Final decision tree to assign the final endemicity of the sub-IU



The tool provides 2 ways of determining the final decision: you can do it manually which is time consuming or you can be assisted by the tool.

3.2.2.2.1.1 Defining the final decision manually

To manually set the final decision for a sub-district, go to the column O (Final decision). Click in the cell and a dropdown list will appear. Select the decision for the sub-district according to the final



decision tree. When you use the manual decision process, the final endemicity is not assigned immediately. To assign the endemicity, click the button “Assign final endemicity”

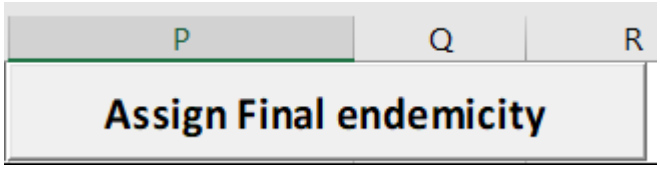
3.2.2.2.1.2 Defining the final decision automatically

To get assisted and save time and energy, you can use the automated decision process. To do so, click the button Suggest final decision at the column O as shown on the above figure. The tool automatically uses the final decision tree to determine the suitable decision and also assign the final endemicity accordingly. However, not all the paths in the decision tree will be explored. Moreover, some decisions may not lead to a proper endemicity in which the value “Unknown” will be assigned to the endemicity. You can modify a suggested decision if it does not fit the reality.

Important note: The automated decision tool process only the rows where both the final decision and the final endemicity cells are blank.

3.2.2.2.2 Assign the final endemicity

Once the decision is made for the sub-IUs, you can now assign the final endemicity. To do so, click the button “Assign Final Endemicity” at the columns P to Q as shown below.



The tool assigns the final endemicity according to the decision made for each sub-IU.

Important note
Every time you click this button, the tool will process all the rows in the worksheet.

4 Summary reports worksheets

The data analysis workbook has 9 pivot-tables that represent 9 summary reports. The reports are selected indicators that demonstrate the importance of the sub-district level implementation and also the estimations of the population requiring treatment and to be considered in the JRSM.

4.1 General settings

The summary reports are built in the same way but describes different indicators. The aggregation levels are the national level, the first level (admin 1), the second level (IU) and the sub-IU level. As they are dynamic pivot tables, you can configure them at your convenience assuming that you have the skills to do so. They have also a filter that use the variable on the inclusion and exclusion criteria. The inclusion/exclusion condition is used to determine the sub-IUs that have their own epidemiological data and then can be included in the comparative advantages of the sub-IU implementation. A sub-IU must comply with two conditions in order to be included in the comparisons. Those conditions are:

- The sub-IU must have a population data
- The sub-IU must have epidemiological data that leads to the calculation of its prevalence and endemicity category
- The sub-IU epidemiological data must have the same characteristics (diagnostic methods, and year) as the district epidemiological data.

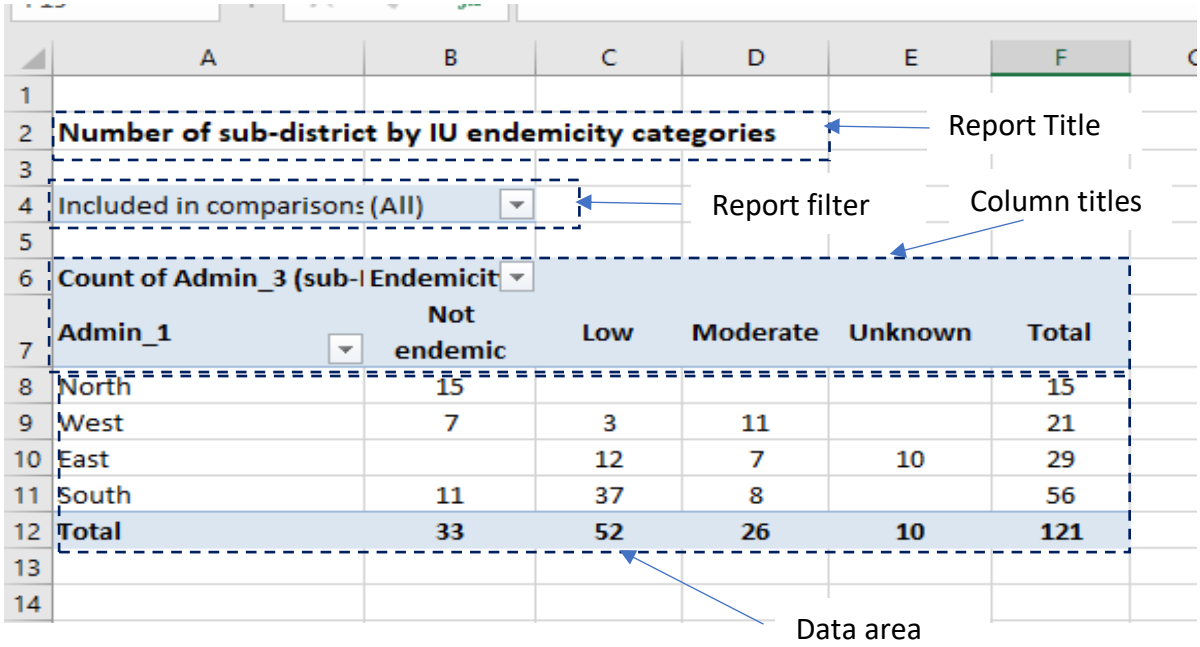
With the above indications, the inclusion/exclusion condition is “Yes” for any sub-IU that comply with the above conditions. You can then use the filter where appropriate to filter the report contents.

The default filter option is “All”. When you turn it to “Yes”, only the sub-IUs that comply with the inclusion conditions will be counted in the summary. The “No” option does not have any particular interest.

The proper use of the filter is further explained in each report section. The figure below is an example of report.

- The reports building blocks are:
- The title,
- The filter (not all the reports have a filter)
- The column description,
- The data summary area.

Figure 4: Summary report template



To change the filter option, click in the dropdown list and select the option you want.

4.2 Filters settings

The table below describe the reports contents according to the filter options.

Table 13: Report filter options and description

Filter option	Report contents
All	Include all sub-IUs
Yes	Include all sub-IUs that have their own data and classified here by their own data
No	Include all sub-IUs that do not have their own data or their own data parameters are not the same as the IU data and classified here by the IU data

4.3 Specific reports

4.3.1 Number of sub-districts by IU endemicity categories

This summary report is built in the worksheet “1_Cat_end”. It shows the number of sub-units by endemicity category calculated with the IU epidemiological data. The figure below is an example of this summary report. This report is aggregated by the admin1 level.

Figure 5: Report on number of sub-districts by IU endemicity categories

	A	B	C	D	E	F
1						
2	**Number of sub-district by IU endemicity categories**					
3						
4	Included in comparisons	(All)				
5						
6	**Count of Admin_3 (sub-I**	**Endemicit**				
7	**Admin_1**	**Not endemic**	**Low**	**Moderate**	**Unknown**	**Total**
8	North	15				15
9	West	7	3	11		21
10	East		12	7	10	29
11	South	11	37	8		56
12	**Total**	**33**	**52**	**26**	**10**	**121**
13						
14						

4.3.2 Number of sub-districts by sub-IU endemicity categories

This summary report is built in the worksheet "1_Cat_end". It shows the number of sub-units by endemicity category calculated with the sub-IUs epidemiological data. It is same as the previous report excel that the endemicity categories are the ones calculated with the sub-IUs data. The figure below is an example of this summary report. This report is aggregated by the admin1 level.

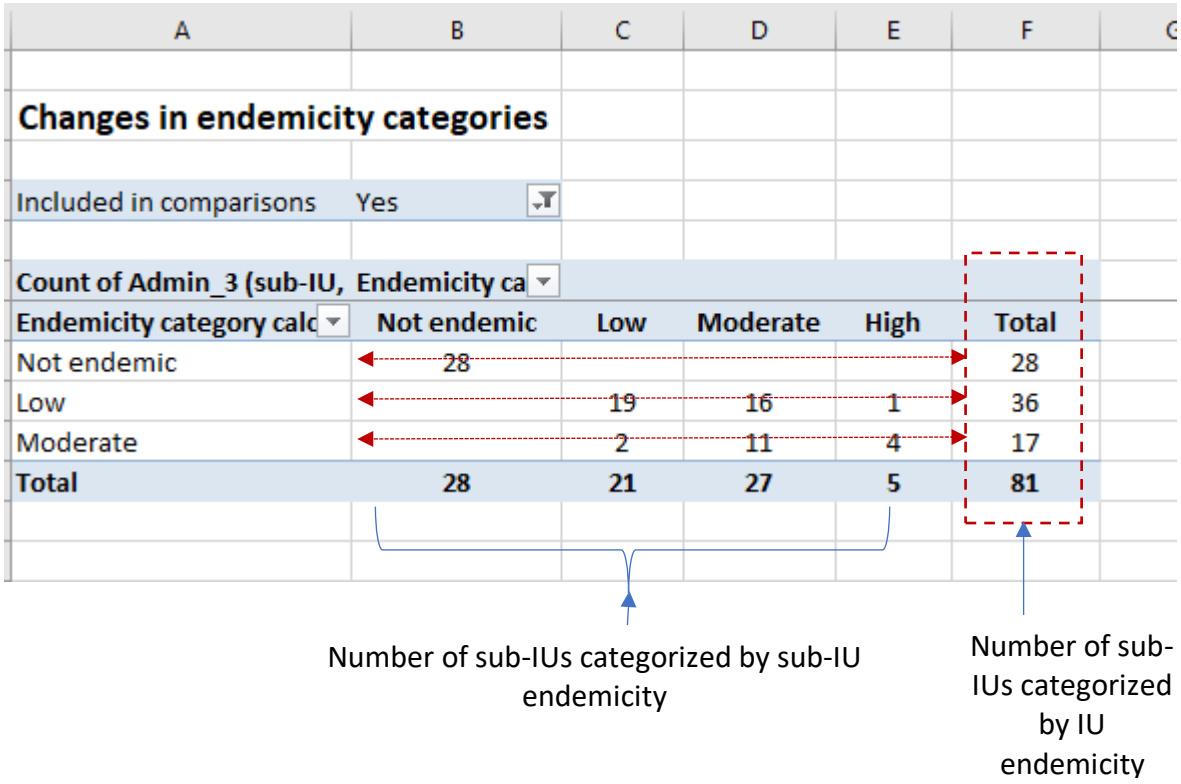
Figure 6: Report on number of sub-districts by sub-IU endemicity categories

	J	K	L	M	N	O	P
	Number of sub-district by sub-IU endemicity categories						
	Included in comparisor	(All)					
	Count of Admin_3 (sut	**Endemicit**					
	Admin_1	**Not endemic**	**Low**	**Moderate**	**High**	**Unknown**	**Total**
	North	14				1	15
	West	5	3	7	4	2	21
	East	1	9	8	1	10	29
	South	12	22	16	2	4	56
	Total	**32**	**34**	**31**	**7**	**17**	**121**

4.3.3 Changes in endemicity categories

The change in endemicity category indicate the state where the endemicity calculated with the IU data may not be the same as the endemicity calculated with the sub-IU data. This report is a contingency table that try to compare the importance of the changes between endemicity categories. This summary report is built in the worksheet "2_Changes". It is aggregated by the national level. It gives a good measure of the changes by turning the filter to "Yes".

Figure 7: Report on the changes in endemicity categories



A	B	C	D	E	F
Changes in endemicity categories					
Included in comparisons	Yes				
Count of Admin_3 (sub-IU,	**Endemicity ca**				
Endemicity category calc	**Not endemic**	**Low**	**Moderate**	**High**	**Total**
Not endemic	28				28
Low		19	16	1	36
Moderate		2	11	4	17
Total	**28**	**21**	**27**	**5**	**81**

In the example above, 28 sub-IUs classified as not endemic using IU data remains the same using sub-IU data. But 17 classified as moderate shift to one as low, 11 as moderate and 4 as high.

4.3.4 Treatment adequacy in district level implementation

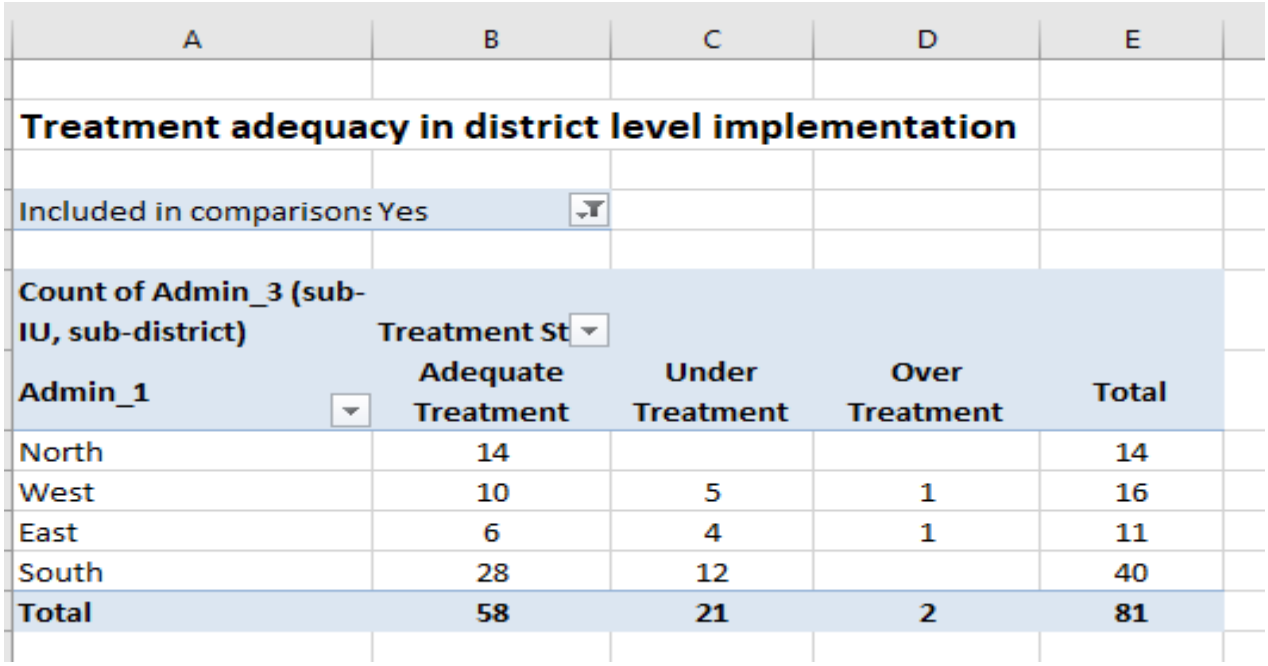
The treatment adequacy report is built on the previous report by transforming the changes in endemicity to treatment adequacy. It built in the worksheet "3_Adequacy". The treatment adequacy tries to highlight the fact that using the IU data for a sub-IUs within the IU instead of using the sub-IU own data can lead to inadequate treatment strategy. The possible scenarios are in the table below.

Table 14: Treatment adequacy definitions

Treatment strategy adequacy	Comments	Changes in endemicity IU data to sub-IU data
Adequate treatment	No change in treatment strategy	Not endemic to not endemic Low to Low Moderate to Moderate High to High
Under treatment	IU data underestimate the sub-IU endemicity	Not endemic to low, moderate, high Low to moderate, high Moderate to high
Over treatment	IU data overestimate the sub-IU endemicity	Low, moderate, high to not endemic High to moderate Moderate to low

The report is aggregated by the national level and gives a good measure of the treatment adequacy by turning the filter to "Yes". The figure below is an example of the treatment strategy adequacy report.

Figure 8: Treatment adequacy in district level implementation



A	B	C	D	E
Treatment adequacy in district level implementation				
Included in comparisons	Yes			
Count of Admin_3 (sub-IU, sub-district)	**Treatment St**			
Admin_1	**Adequate Treatment**	**Under Treatment**	**Over Treatment**	**Total**
North	14			14
West	10	5	1	16
East	6	4	1	11
South	28	12		40
Total	**58**	**21**	**2**	**81**

In the example below, 58 sub-units are adequately treated, while 21 are under treated and 2 over treated.

4.3.5 Population estimations in district level compared to sub-District level implementation

This summary report is to compare population requiring PC using IU and sub-IU endemicities. It relevant only in the case that the sub-IU has its own data. The filter must be turned to "Yes". The estimation is done on a yearly average since treatment do not occur every year every time. The table below shows the ratios applied to each population group and endemicity category conformally to the WHO guidelines.

Table 15: Ratios used in the calculation of the populations

Endemicity category	SAC	Adults
Not endemic	0	0
Low	0.33 (every 3 years)	0
Moderate	0.5 (every 2 years)	0.2
High	1 (every year)	1

The figure below shows an example of this report.

Figure 9: Report on the comparison of population estimations

Population estimations in district level compared to sub-District level implementation

Included in comparisons: Yes

Admin_1	Number of sub-Ius with epidemiological data	Sum of SAC estimates (IU endemicity)	Sum of Adults estimates (IU endemicity)	Sum of SAC estimates (sub-IU endemicity)	Sum of Adults estimates (sub-IU endemicity)
North	14	0	0	0	0
West	16	59,397	45,097	85,693	42,810
East	11	71,080	25,132	80,914	24,978
South	40	186,252	21,803	207,743	77,360
Total	**81**	**316,729**	**92,032**	**374,350**	**145,148**

4.3.6 Drugs estimations in district level compared to sub-district level implementation

The PZQ estimations reports is built in worksheet "5_PZQ_req_d_sd" in the same way as the previous report. The required quantities of tablets are calculated by multiplying the population estimation by 2.5 for the SAC and 3 for the adults. The figure below shows an example of this report.

Figure 10: Report on the comparison of drug estimates

Drugs estimations in district level compared to sub-district level implementation

Included in comparisons: Yes

Admin_1	Number of sub-Ius with epidemiological data	Sum of Drugs estimates for SAC (IU endemicity)	Sum of Drugs estimates for adults (IU endemicity)	Sum of Drugs estimates for SAC (sub-IU endemicity)	Sum of Drugs estimates for adults (sub-IU endemicity)
North	14	0	0	0	0
West	16	148,494	135,291	214,234	128,430
East	11	177,702	75,396	202,286	74,934
South	40	465,639	65,409	519,366	232,080
Total	**81**	**791,835**	**276,096**	**935,886**	**435,444**

4.3.7 Inequity in the IU level implementation

This report is a summary of the 3 previous reports: Treatment strategy adequacy, population requiring PC and drug needs to show the benefits of the sub-IU implementation. It shows:

- The number of SAC adequately treated
- The number of SAC missing treatment (not treated because of the underestimation of the sub-IU prevalence)

- The number of SAC unnecessary treated (because of the overestimation of the sub-IU prevalence)
- The amount unclaimed drugs (underestimation of drugs consecutive to endemicity underestimate)
- The amount drugs misused (over treatment)

Those indicators show the counter-performance of the district level implementation by leaving people behind while dispensing resources where not needed, all in contradiction with the principle "One Health for All".

The figure below shows an example of this report.

Figure 11: Report on the indictors to measure the inequity in the district level implementation

A	B	C	D	E	F	G
Treatments inequity in district level implementation						
Included in comparisons	Yes					
Admin 1	**Number of sub-Ius with epidemiological data**	**Sum of SAC adequately treated**	**Sum of SAC missing treatment**	**Sum of SAC unecessary treated**	**Sum of Unclaimed drugs**	**Sum of Drugs misused**
North	14	0	0	0	0	0
West	16	25,946	29,548	3,252	73,871	8,130
East	11	47,724	12,130	2,296	30,326	5,740
South	40	147,524	21,491	0	53,731	0
Grand Total	**81**	**221,194**	**63,169**	**5,548**	**157,928**	**13,870**

4.3.8 Population of SAC and adults requiring PC by sub-implementation unit

This report shows the population requiring PC for the sub-IU after all considerations that permit to determine its final endemicity category. The population data in this report are the final to be used to fill in the JRSM. Since the JRSM is by IU, the next report is the aggregation by IU. No ratios are applied in the calculation since the treatment will cover the whole target population. Only the frequency of treatment will change according to the endemicity category. The figure below shows an example of this report.

Figure 12: Population requiring PC by sub-IU

A	B	C	D	E	F
Population of SAC and adults requiring PC by sub-implementation unit					
Admin_1	**Admin_2 (IU)**	**Admin_3 (sub-IU, sub-district)**	**Year of population**	**SAC living in sub-IU requiring PC**	**Adults living in sub-IU requiring PC**
West	Center-West	Ward-016	2019	0	0
		Ward-017	2019	0	0
		Ward-018	2019	0	0
		Ward-019	2019	0	0
		Ward-020	2019	0	0
		Ward-021	2019	0	0
		Ward-022	2019	0	0
		Ward-023	2019	29006	79452
		Ward-024	2019	18284	50082
		Ward-025	2019	7970	21830
		Ward-026	2019	10284	28170
		Ward-027	2019	2706	7412
		Ward-028	2019	3324	9104
		Ward-029	2019	22799	62449
		Ward-030	2019	17157	46995
	Urban West	Ward-031	2019	19127	52391
		Ward-032	2019	6276	17190
		Ward-033	2019	5026	13766
		Ward-034	2019	27287	74743
		Ward-035	2019	200	548
		Ward-036	2019	9030	24734
East	East Rural Council	Ward-037	2019	0	0
		Ward-038	2019	0	0
		Ward-039	2019	0	0
		Ward-040	2019	0	0
		Ward-041	2019	0	0
		Ward-042	2019	0	0
		Ward-043	2019	0	0
		Ward-044	2019	0	0
		Ward-045	2019	0	0
		Ward-046	2019	0	0
		Ward-047	2019	10012	27424
		Ward-048	2019	1523	4171
	East Town Council	Ward-049	2019	14748	40396
		Ward-050	2019	1327	3635

4.3.9 Population of SAC and adults requiring PC by implementation unit

This report is the aggregation of the previous report by IU level. The figure 12 below is an example of this report.

Figure 13: Population requiring PC by IU and by age

A	B	C	D	E
Population of SAC and adults requiring PC by implementation unit				
Admin_1	**Admin_2 (IU)**	**Year of population**	**SAC living in IU requiring PC**	**Adults living in IU requiring PC**
North	**North-East**	2019	0	0
	North-West	2019	0	0
West	**Center-West**	2019	111,530	305,494
	Urban West	2019	66,946	183,372
East	**East Rural Council**	2019	11,535	31,595
	East Town Council	2019	57,326	157,024
	East-Forest	2019	191,646	524,942
South	**South-East**	2019	0	0
	South-West	2019	125,329	343,291
	Middle South	2019	103,069	282,321
	Far South	2019	318,530	872,498
	South-Savannah	2019	73,393	201,033
Total			**1,059,304**	**2,901,570**