

OPTIMIZING SCHISTOSOMIASIS MDA IMPLEMENTATION IN COUNTRIES

Data Analysis Tools

June-July 2019

Based on joint application packages (JAP) received from countries, preventive chemotherapy for schistosomiasis using praziquantel distribution is currently largely implemented at district levels. This is in a wider scale than may be needed, since schistosomiasis transmission is focalized 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. The current work consists of analysis of selected countries where review of sub-district level datasets may result in adjusting implementation to lower levels than is currently targeted

Table of Contents

- 1. Software requirements 2
- 2. Data requirements 2
- 3. Data Structure..... 2
- 4. Data Quality Control..... 2
 - Data Completeness* 2
 - Data Integrity* 2
 - Data accuracy and suitability*..... 3
 - Serialization* 4
- 5. Calculations 4
- 6. Annexes..... 7

1. Software requirements

The analysis will be carried out using Microsoft Excel 2010 and later release. The latest release is 2016.

2. Data requirements

See the document entitled “Data Analysis Requirements”

3. Data Structure

The complete analysis will be performed in one Excel workbook that allows macro. The extension of such workbook is xlsx. Three worksheets will be created within the workbook.

- The worksheet for the epidemiological data will be named “EpiData”.
- The worksheet for the demographic data will be named “DemoData”.
- A third worksheet will be created to combine the epidemiological and demographic data which will be complemented by the output data (data derived from the input data by appropriate formula and that are needed for the purpose of the sub district level implementation).

4. Data Quality Control

The data quality assessment is the first step of the analysis. It will help to judge the fitness of the data to the purpose of the sub district analysis. This quality assessment will include essential criteria and parameters that are commonly reviewed in data quality assessment.

Notes: The quality assessment at this level will exclude any parameters that have affected the data during its collection. It will only take into account the data as it is presented.

The quality assessment will include completeness, Integrity, Consistency, Cleanliness.

Data Completeness

Listing of the sub-districts. The list of the sub-districts must be complete. If there is any sub-district that are not listed, it should be noted and document so that the final results are interpreted taking into account the incompleteness of the sub-districts. If their number is known, they will be included in the sub-district list with a temporary name that will be replaced later with the correct name. The completeness of the sub-district list is to be assessed in the demographic data.

Data Integrity

The data integrity will ensure that the same data element has the same attributes and meaning throughout. As an example, the name of a district must be written in the same way in the epidemiological and demographic data.

- Naming convention

All data elements that are in both the epidemiological and demographic data will be checked and harmonised if needed. Name of places (Country, Province, Region, State, District, IU, Sub-district, Site, School, Locality, Village, etc.) will be harmonised and made unique. The same name cannot be used for 2 different places within the same health administrative level. If same names are used across different administrative levels, they will be differentiated.

Example: Two sub-districts have the name DJEGBE in two different IUs Glazoué and Abomey

Before

Country	Province	District	Sub district
Benin	Collines	Glazoué	DJEGBE
Benin	Zou	Abomey	DJEGBE

After

Country	Province	District	Sub district
Benin	Collines	Glazoué	DJEGBE (Glazoué)
Benin	Zou	Abomey	DJEGBE (Abomey)

The IU names are added in brackets to the initial name DJEGBE.

- Codification

Whenever possible, codes can be assigned to the IU names and the codes will be used instead of the names to perform some operations and calculations that require the creation of a link between the datasets.

Data accuracy and suitability

The available site data were collected under different circumstances and using different methodologies. Thus, they may not have all the same level of accuracy, certainty, precision, etc. Moreover, a single site may have more than a record all trying to explain the same information. For example, there may be a site surveyed twice, first time using parasitological test and second time using clinical test. Regardless the results of both tests, the parasitological test is far preferred to the clinical test. Another example is 2 different surveys in the same site using both same parasitological test but carried out one in 2000 and another in 2015. The most recent test will probably be preferred.

In this context the record will be graded first according to the diagnostic methods and then to the age of the data. The table below shows the grading from 1 (most suitable) to 5 (less suitable) the diagnostic methods.

Diagnostic Type	Diagnostic Methods	Grade
Parasitological	Urine Filtration	1
	Kato-Katz	1
	Urine Sedimentation	1
	Other microscopy	2
Clinical	BIU	3
	Dipstick	3
	Urine Other	4
Serological	CCA	4
Other	Unknown	5
	not specified	5
	Other	5
	Other molecular	5

Following this grading system, a site that have more than one record will be used in the calculation only once with its most suitable data.

Serialization

It consists of attributing a unique serial number to each row in the worksheet. This will be just an incremental whole number from 1 to the total number of rows in the worksheet. This will help to have a unique ID for each data row.

5. Calculations

Various operations will be performed to calculate the output indicators data. All calculations are performed for the sub-district by considering the 2 situations (1) the sub-district prevalence calculated for the district and (2) the sub-district prevalence calculated for the sub-district.

1. Site prevalence - SP

The site prevalence if not already calculated is calculated as below:

$$SP = (\text{Number of people positives}) / (\text{Number of people examined}) * 100$$

2. District or Sub-district Prevalence

The district or sub-district prevalence is calculated as the mean prevalence within the district or the sub-district.

$$DP = (\text{Total positives}) / (\text{Total examined}) * 100$$

$$SDP = (\text{Total positives}) / (\text{Total examined}) * 100$$

3. Prevalence Range - PR

The prevalence range inform about the lowest and the highest site prevalence within the district or the sub-district

PR = [Min site prevalence – Max site prevalence]

4. Endemicity category

The endemicity category is determined as recommended in the WHO guidelines. For more information refer to the annex 1. It depends on the ecological zone determination (using GIS technology or local knowledge of the area), and the prevalence value (average, maximum, quantiles, etc.) and the diagnostic methods (parasitological or clinical).

- Choice of the prevalence value

The endemicity category will be determined by the prevalence value within the district or sub-district. The prevalence value to be used within the district is still not yet well defined. Most of times, the average value is recommended. But in few cases, the maximum value is used. Some countries have preferred the use of the upper limit (95%CI) of the average value. The choice of the prevalence value to be used may depend on some other considerations such as resources, size of the mapping unit. The maximum can work for a small mapping unit or area (district or sub-district) and average will be much more indicated for wider units.

For this workshop, we will use the average value of the prevalence within the district or sub-district unless otherwise desired by the country based on the ongoing implementation.

- Choice of the diagnostic methods

In some cases more than one diagnostic method has been used. For example, questionnaire for history of haematuria (clinical method) may be used in conjunction with urine filtration (parasitological method) for urogenital schistosomiasis. For intestinal schistosomiasis, Kato-Katz has been widely used, while some countries also collected CCA (serological test) data.

As the prevalence thresholds are not the same for the different methods, it is not recommended to do any prevalence estimation by combining the different tests results. Where the thresholds are the same (parasitological methods), the combined prevalence should be calculated from the raw individual datasets.

If the combined prevalence is performed from the individual dataset, the categorisation is performed using the combined prevalence (Kato-Katz for intestinal species and urine filtration for *S. haematobium*).

If only the specific diagnostic and species prevalence are available, then the categorisation will be performed in each case and the highest prevalence category will be assigned to the district or sub-district.

The figure in annex 2 describe the different configurations

1. Population estimates

The population to be treated is determined by the endemicity category as recommended in the WHO guidelines.

2. Drugs estimates

The amount of drugs needed is calculated by multiplying the population to be treated by a factor that depends on the age. This factor is 2.5 for school age children and 3 for adults.

3. Treatment Strategy Adequacy

The treatment adequacy is determined by comparing the 2 treatment strategies of a sub-district respectively to the district prevalence and the sub-district prevalence. It leads to 3 parameters for which drugs will be estimated.

- No change in strategy: the treatment strategy does not change from district implementation to sub-district implementation
- Under treatment: the treatment strategy in sub-district implementation expose more people than in district implementation
- Over treatment: the treatment strategy in sub-district implementation expose less people than in district implementation

In each category, drugs amounts will be estimated.

6. Annexes

Annex 1: Updated recommended treatment strategy for controlling morbidity due to schistosomiasis including CCA prevalence (interim guidance)

Category	Baseline prevalence among school-age children	Action to be taken		Additional interventions
High-risk community	<p>$\geq 50\%$ by parasitological methods (intestinal and urogenital schistosomiasis)</p> <p>or</p> <p>$\geq 30\%$ if based on questionnaires for history of haematuria</p> <p>or</p> <p>$\geq 60\%$ by CCA in <i>S. mansoni</i> endemic areas</p>	Treat all school-age children (enrolled and not enrolled) once a year	Also treat adults considered to be at risk (from special groups to entire communities living in endemic areas)	Water, sanitation Improve, hygiene education (WASH) Snail control
Moderate-risk community	<p>$\geq 10\%$ but $< 50\%$ by parasitological methods (intestinal and urogenital schistosomiasis)</p> <p>Or</p> <p>$< 30\%$ by questionnaire for history of haematuria</p> <p>Or</p> <p>$\geq 15\%$ but $< 60\%$ by CCA in <i>S. mansoni</i> endemic areas</p>	Treat all school-age children (enrolled and not enrolled) once every 2 years (essentially treat 50% of this age group each year)	Also treat adults considered to be at risk	Water, sanitation and hygiene education (WASH) Snail control
Low-risk community	<p>$< 10\%$ by parasitological methods (intestinal and urogenital schistosomiasis)</p> <p>Or</p> <p>$< 15\%$ by CCA in <i>S. mansoni</i> endemic areas</p>	Treat all school-age children (enrolled and not enrolled) twice during their primary schooling age (treat at least 33% of this age group each year)	Praziquantel should be available in dispensaries and clinics for treatment of symptomatic cases	Water, sanitation and hygiene education (WASH) Snail control

Annex 2: Endemicity classification configurations

