# Standard Operating Procedure for MSF Antibiotic Consumption Analysis

This document describes how to use templates for consumption analysis based on HIS and DHIS2 data combined with consumption tools in specific projects (in the context of antibiotic usage surveillance). The SOP describes project structure (where to find what), how to update code, add new data and new countries/projects.

Contents

[Standard Operating Procedure for MSF OCA Time Series Analysis 1](#_Toc77786336)

[2 Project Structure 1](#_Toc77786337)

[2.1 Location 1](#_Toc77786338)

[2.2 Layout 1](#_Toc77786339)

[2.3 Important Files 1](#_Toc77786340)

[2.4 Important Functions 2](#_Toc77786341)

[3 Producing Graphs 2](#_Toc77786342)

[3.1 Preparing Data 2](#_Toc77786343)

[3.2 Preparing Code 3](#_Toc77786344)

[3.3 Running Code 5](#_Toc77786345)

[4 Adding Additional Projects 6](#_Toc77786346)

## 2 Project Structure

### 2.1 Location

(Ideally the project will be moved to SharePoint shortly)

### 2.2 Layout

The project is set up as a R workbook, meaning you only need to open one file. There are two folders labelled “IPD Data” and “Pharmacy Data”.

The “IPD Data” folder contains two sub-folders per project, i.e., “Helmand\_DHIS” and “Helmand\_OLD”, where the DHIS folder contains the file “events.csv” imported from DHIS for the specific project. The “Helmand\_OLD” folder contains the processed output from the project located on SharePoint (link redacted), containing the IPDmed file for each year.

The “Pharmacy Data” folder contains one sub-folder per project, i.e., “Helmand”. This sub-folder contains at most one consumption tool per year as an .xlsx file.

### 2.3 Important Files

The main directory contains two files: “Main\_Scripts.Rmd”, and “DDDs.txt”. The first file is the body of code, which needs to be opened in an R environment. The second file is an (editable) list of drugs, their defined daily dosage, and their AWARE classification according to the WHO (access/watch/reserve).

### 2.4 Important Functions

Each block of code in the workbook contains (generally) one function with a detailed description of the inputs, outputs, and how the function itself works.

## 3 Producing Graphs

### 3.1 Preparing Data

For the code to run, the appropriate data needs to be placed in the correct folders and named appropriately. The data needs to be placed in the folders as described in [2.2](#_2.2_Layout).

#### 3.1.1 Old IPD files

The processed inpatient data files should be in the form “\*.med”. These files can be obtained from the SharePoint link in [2.2](#_2.2_Layout). They should have a name like “IPDmed v1.2.6.2\_Bangladesh KTP BHC MSF-OCA 6-1-2018.med”. Once the files are downloaded, they can be renamed at will. Make sure that they are all placed in the folder “/IPD Data/Location\_OLD/”.

#### 3.1.2 DHIS2 data

DHIS2 data can be downloaded from the MSF OCA HIS portal. First, click on “Event reports”, choose “Events” (instead of “Aggregated Values”), set the Program to “Inpatient admissions” then select the following:

1. “Admission ward”,
2. “Age (combined, in years)”,
3. “Age (years)”,
4. “Date of admission”,
5. “Date of exit”,
6. “Diagnosis at exit – primary”.

More categories can be selected, but it will only increase the file size and slightly slow down processing. Also, make sure in the “Selected data items” tab, the “Age (combined, in years)” item has a field containing “Age all years categorised”. This needs to be changed to “No range set”. The same must be done for “Age (years)”.

Next, set the time period to the maximum amount- either by setting the start and end dates, or by selecting “Yearly” and pressing the double arrow “>>” to select all years available.

Then, in the “Organisation units” tab, select the location, then the project, then the inpatient services file (it may have a different name or be in another sub-folder) as shown in Figure 1.

Finally, press “Update” – this should display a table containing the information selected as well as a column for “Event date” and for “Organisation unit”. Now, click “Download” -> “CSV”.

After downloading, the data should be stored in the appropriate folder in processed as described in [2.2](#_2.2_Layout). The file *must* be named “events.csv” and *must* be located in “/IPD Data/Location\_DHIS/”.

#### 3.1.3 Consumption tools

The consumption tools must be in an “.xlsx” format. Originally the tools are saved as “.xlsb” files so they must be opened and saved as “.xlsx” format. This creates a copy of the consumption tool which the program can use. Also, they must have the year (20xx) anywhere in the file name. The number of consumption tools required varies between projects – for instance, Helmand only required two (the latest 2021 tool and the older 2017 tool) to cover all cases, whilst others require one every two years. The easy way to tell whether there is missing data is to look for areas of zeros in the graphs produced. To obtain the consumption tools, contact Kristina (link redacted) and ask for the consumption tools for the project of interest.

Graphical user interface, text, application

Description automatically generated

Figure : Selecting the correct Organisation units

### 3.2 Preparing Code

All the code is in the “Main\_Scripts.Rmd” file as mentioned above. Within this file, there is one block of code which contains everything required to run the scripts. To find the correct block, first press ‘Alt + o’ to collapse all code chunks, then scroll down until you find the block labelled ‘{r MAIN CONTROL}’. Then, press the button to run all code chunks above, as shown in Figure 2.

Once this has been done, press the  symbol immediately to the left of the code block to open the main control block. As seen in Figure 3, this contains options for customisation/modification.

The most important options are:

1. The start date, currently set to 1 Jan 2018. This determines whether files are read or skipped based on the year, as well as when to start plotting from.
2. The include/exclude toggles and wards for the DHIS data. These determine whether to filter by ward, and which wards to include/exclude.
3. The include departments for the pharmacy data. This must include all departments (excluding paediatric) on which consumption is to be analysed.

One key thing to note is that each project may have different names for each department, and these names may be different year by year. To find out if a newly added project has a different naming convention, set the variable ‘broadcast\_departments’ to TRUE, then run the code for the new project. The scripts will then display all wards found for each consumption tool. Once this is done, add any departments desired into the ‘include\_departments’ list, set ‘broadcast\_departments’ to FALSE, and re-run the code. Now the additional departments will be included in the output.

The final step before running the code is to make sure to specify what graphs to produce. Currently, the options are:

1. DDDs/1k Bed Days plot for a specific drug
2. DDDs/1k Bed Days plot for the top N drugs, where N is an integer
3. Quarterly Access/Watch/Reserve plots (in DDDs/1k Bed Days)

These are shown in Figure 4. Also shown is how to generate the processed data for a specific project – this must be replicated for any additional projects of interest. Note that the variable names (e.g., ‘pharmacy\_Helmand’) must be different for a different project. The easiest way to add additional projects is to copy-paste an existing block such as in Figure 4 and replace all instances of ‘Helmand’ with the new project name.

Graphical user interface, application

Description automatically generated

Figure : Running all code above Main Control

Text

Description automatically generatedFigure : Customisation options

Text

Description automatically generatedFigure : Generating plots

### 3.3 Running Code

Once the code has been prepared as described above, it is ready to be used. Make sure that any projects that you do not wish to process are commented out (a *#* at the start of the line, denoted by any following code turning green). Then, press ‘Ctrl + Shift + Enter’ (or press the  at the top right of the ‘MAIN CONTROL’ code block) to run it. Note that it may take up to 10 seconds per consumption tool per project, depending on the size.

Once this has finished running, all the plots will have been generated. You can view a specific plot by typing the name of the plot into the R console. If you want to change a plot (such as changing the time period to plot, or the number of drugs, or the specific drug), then instead of re-running the entire code chunk, make your changes to the line of code for that plot then press ‘Ctrl + Enter’ to only run that one line. This is much faster than running the entire block again as the data has already been read and processed.

After typing the name of the plot into the R console, the plot should appear on the right-hand side of the R window. This can then be right-clicked and either copied or saved. Alternatively, just above the plot there is a button which says ‘Export’. Click on this and choose from the options to save as an image or PDF or copy to clipboard.

## 4 Adding Additional Projects

In order to run the code on a new project, the following steps must be followed:

1. Set up folders as described in [2.2](#_2.2_Layout).
2. Acquire the relevant data as described in [3.1](#_3.1_Preparing_Data) and place into the correct folders.
3. Add additional lines of code to the MAIN CONTROL chunk for the new project (copy Figure 4).
4. Set the variable ‘broadcast\_departments’ to TRUE.
5. If there are departments printed which you think should be included but are not currently in the list ‘include\_departments’, add them to it. Reset ‘broadcast\_departments’ to FALSE.
6. Run the code as described in [3.3](#_3.3_Running_Code).