

In Vitro Module: Data Management and Statistical Analysis Plan  
(DMSAP)  
Version 1.1

In Vitro Module  
WorldWide Antimalarial Resistance Network (WWARN)



Suggested citation: In Vitro Module, WWARN, 2013. Data Management and Statistical Analysis Plan.

### Version History

Version number	Revision(s) & reason for amendment	Release date
v1.0		
V1.1	<p>Separation of core criteria and range warnings so that range warning does not lead to removal from core</p> <p>Modification of range warnings to produce Range Low and Range High warnings</p> <p>Changes in the Data set submission process and Data cleaning process to include the use of IVART Online by the investigator</p>	08/03/2013

For more information, contact:

[invitro@wwarn.org](mailto:invitro@wwarn.org)

WorldWide Antimalarial Resistance Network (WWARN)

[www.wwarn.org](http://www.wwarn.org)

## Contents

1. Scope.....	4
2. Introduction .....	4
2.1 Generating evidence of antimalarial drug efficacy .....	4
2.2 WWARN In Vitro Data Processing .....	4
3. Dataset submission process.....	5
3.1 Data submission steps.....	5
3.2 Study files .....	6
3.3 Sample data files .....	6
3.4 Primary (raw) data files .....	6
3.5 Criteria for acceptance.....	7
3.6 Storage .....	7
4. Data Curation.....	7
4.1 Curation by WWARN .....	7
4.2 Curation by the investigator.....	7
5. Analysis via the In Vitro Analysis and Reporting Tool (IVART) .....	7
5.1 Sigmoid $E_{max}$ model .....	8
5.2 Normalisation.....	8
5.3 Non-linear regression.....	9
5.4 Core Criteria for summary analyses .....	9
5.5 Range warnings .....	10
6. Reporting.....	10
6.1 Study report .....	10
6.2 Microsoft Excel™ files.....	11
6.3 Review by curator and contributor .....	11
7. WWARN Explorer .....	11
8. Concluding remarks .....	11
9. References .....	12

# 1. Scope

A key aim of the Worldwide Antimalarial Resistance Network (WWARN) is to monitor geospatial and temporal trends in antimalarial drug resistance. The purpose of the In Vitro Data Management and Statistical Analytical Plan (DMSAP) is to provide a clear and transparent description of the methodology by which WWARN handles and analyses *in vitro* data. *In vitro* susceptibility data need to be collated at an individual assay level so that standardized methodologies and definitions can be applied. Documenting the entire process by which data are uploaded, transformed, analysed and presented ensures reproducibility. It also provides a framework for discussing and developing such methodologies.

## 2. Introduction

*In vitro* drug susceptibility remains a central pillar of antimalarial efficacy surveillance. The assessment of antimalarial drug susceptibility in parasites straight from patients (*ex vivo* parasites that are not culture-adapted) is largely independent of clinical factors and hence provides information that complements clinical assessment of drug efficacy. The WWARN In Vitro Module aims to enhance the amount of *in vitro* data available, increase scientific capacity in endemic countries and provide intelligence on resistance or susceptibility to various antimalarials.

The In Vitro Module is working towards these goals in the following ways:

- Facilitating the inclusion in a Data Repository of results from *in vitro* susceptibility studies contributed by research groups, Non-Governmental Organisations (NGOs) or National Malaria Control Programmes (NMCPs) from around the world.
- Developing high-throughput data management tools to facilitate standardised analysis of large datasets.
- Promoting the use of validated antimalarials and reference strains to enhance comparability of results between heterogeneous studies.

### 2.1 Generating evidence of antimalarial drug efficacy

Determination of the *in vitro* susceptibility of field *P. falciparum* isolates to antimalarial drugs is undertaken via a variety of measurement systems and analytical approaches. These, as well as variation in culture methods, have developed according to specific research questions, characteristics of the study site or simple logistical constraints. Such variations in design, methodology, analysis and presentation of *in vitro* studies constitute a challenge for the collation of data from different centres. WWARN's approach to this issue, across all modules, is to work with complete sets of primary (raw) data, allowing the characteristics of the methodology to be understood, and analyses to be undertaken via a standardised approach. For *in vitro* studies, the primary data are the raw output from *ex vivo* assessment of drug effects on an individual isolate for a single drug. In almost all cases, the dataset for an isolate comprises many such individual assessments for each of several different drugs.

### 2.2 WWARN In Vitro Data Processing

A series of steps will be followed for each dataset submitted to WWARN:

- I. Submission of primary experimental and sample data from an *in vitro* study

- II. Transformation and Data analysis applying the In Vitro Analysis and Reporting Tool (IVART), uniform analytical methodology to provide reproducible estimates of drug susceptibility
- III. Reporting of results to the contributor
- IV. Agreement of the contributor that the contributed data can be included in pooled analyses and/or presented on the WWARN Explorer
- V. Display of results on the WWARN Explorer, an interactive, online tool which allows users to perform custom queries of data from multiple studies and modules and visualise the results using dynamic mapping

WWARN notes that individual researchers may take different approaches to data management and analysis. It is inevitable that WWARN-derived drug susceptibility results will vary to some extent compared to analyses performed by the data contributor. Any differences do not reflect a value judgment as to which analytical approach is correct. A uniform approach is taken only to apply standardised methodologies and minimise bias on geospatial and temporal trends derived from the many studies compiled in the Data Repository.

The purpose of this document is to describe the key processes of data management and analysis that will be undertaken on individual in vitro datasets by WWARN.

### 3. Dataset submission process

Datasets can be submitted to WWARN in two different ways:

1. Using the WWARN online submission system in which a WWARN curator will clean and analyse the data using IVART.
2. Investigator can do the data curation and analysis themselves and submit the data set to WWARN using the IVART Online web-application.

Data contributors must accept the Terms of Submission [available at <http://www.wwarn.org/data/usage>] during the submission process.

#### 3.1 Data submission steps

The data submission steps are:

- I. Study Registration: contributors enter their study title and each study is assigned a unique WWARN identifier. Contributors can also tick one or more boxes to indicate the type of data submitted (clinical, molecular, *in vitro* and/or pharmacology).
- II. Permissions: each study may have any number of administrators, assigned by the data contributor who created the study title and accepted the Terms of Submission. An administrator can access the study, upload files and edit supplied information.
- III. Files: contributors are asked to submit data files and supporting documentation including a data dictionary, protocols and publications.
- IV. Publications: contributors can provide the Pub Med ID, citation or DOI for publications relating to the submitted data.
- V. Acknowledgements: the names of acknowledged individuals and institutions will appear in the study details displayed in WWARN Explorer.
- VI. Study information: data contributors are asked to provide information on the study site, laboratory conditions and drug sources in the study site questionnaire. They may

enter this information themselves or provide relevant protocols and publications which will be used by WWARN Data Managers to extract relevant data. If such information is common to all files uploaded by one group, the information can be saved as a unique protocol, and referred to in subsequent submissions.

### 3.2 Study files

WWARN recognises that contributors organize their data in a variety of different ways and therefore accepts most types and numbers of data file. Preferably, a contributor may upload a data file listing information about each sample and a primary data file containing the output from individual *in vitro* assays in the 96-plate format. If using IVART Online, data must be uploaded in the 96-plate format.

### 3.3 Sample data files

These typically contain one set of data for each isolate:

- Unique identifiers for the sample
- Whether the sample is a clinical isolate (*ex vivo*) or a laboratory reference strain
- Patient variables, if available: age, sex, haematocrit/haemoglobin level, white cell count, parasitaemia (per 200 WBC, 1000 RBC or microlitre)
- Date and time the parasite sample was taken
- Time when culture commenced (allows calculation of delay)
- Location where sample taken (or where malaria acquired in travellers)
- Assay method used
- Duration of assay

### 3.4 Primary (raw) data files

These files contain primary experimental data values from *ex vivo* assays, along with identifying information to allow linking to the sample file. Complete primary datasets, including data from apparently 'failed' assays, should be submitted so that the most accurate picture of a study can be developed. If 96-well plate data are provided, the contributor should provide a description of the plate layout in terms of drug and concentration. Having provided this once, further datasets using the same layout can be transformed using the same template.

An example of the information anticipated is shown in Figure 1.

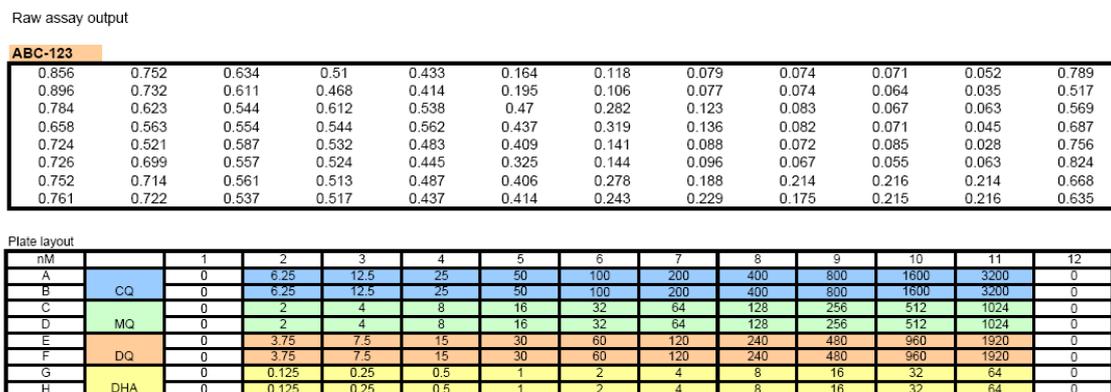


Figure 1: Example of primary (raw) data

### 3.5 Criteria for acceptance

It will not be possible to process data sets without the primary output data, unique identifiers, geographical sample/acquisition location, dates of sample collection and drug-free control wells. For each drug, assay data from a minimum of ten individual isolates should be submitted.

### 3.6 Storage

The new dataset will be recorded as having been received for curation and further progress through the curation and analysis process will be communicated to the contributor. Data will be stored on the WWARN server ensuring a high level of security. Individual data files will be accessible only to the original contributors, colleagues that they authorize and selected WWARN staff.

In order to ensure compliance with its legal and ethical obligations, WWARN requires that all data has been anonymised prior to submission.

## 4. Data Curation

Curation of submitted data involves transformation of drug, concentration and readout values, along with relevant sample data, into a standardised format, allowing single study analysis, the generation of a study report and, with permission, visualisation of summary study information on WWARN Explorer. The curated data is stored in a Data Repository for future pooling with other datasets and meta-analyses.

### 4.1 Curation by WWARN

If raw data is submitted to WWARN through the online submission system, a WWARN curator will interact with the contributor so the data is clearly understood and correctly transformed and analysed. It is critical that there is a clear understanding in terms of drug names and concentrations (nM concentrations are standard), unique identifiers (matched between sample and primary data files) and place names and locations.

### 4.2 Curation by the investigator

To use IVART Online, data in the 96-well format needs to go through the process of "tagging" by the investigator, for the data to be annotated and recognised by IVART. After upload, the tagged data file goes through an automatic checking function in IVART Online, reporting duplicates, missing data, date errors and other inconsistencies that need to be resolved prior to data analysis. Details about using IVART Online can be found in the [Step by Step User Guide](#).

## 5. Analysis via the In Vitro Analysis and Reporting Tool (IVART)

Concentration-effect data will be analysed by non-linear regression. Several non-linear regression models are broadly appropriate for this purpose and have been used in published work to calculate IC<sub>50</sub> values and other variables from *in vitro* data. Results in the WWARN In

Vitro Study Report are derived from analysis based on a 4-parameter sigmoid  $E_{max}$  model, with fixed upper and lower constraints according to control values (i.e. a 2-parameter model).

### 5.1 Sigmoid $E_{max}$ model

A symmetrical sigmoidal curve is produced with an upper and a lower constraint.

- The upper constraint is defined by the value  $E(C_0)$ , defined as the mean effect across all wells on a plate which contain no drug.
- The lower constraint is defined by the value  $E_{min}$ , defined as the average effect over the two concentrations with the lowest mean effect for a particular drug.

An example of data with these two constraints is shown in Figure 2.

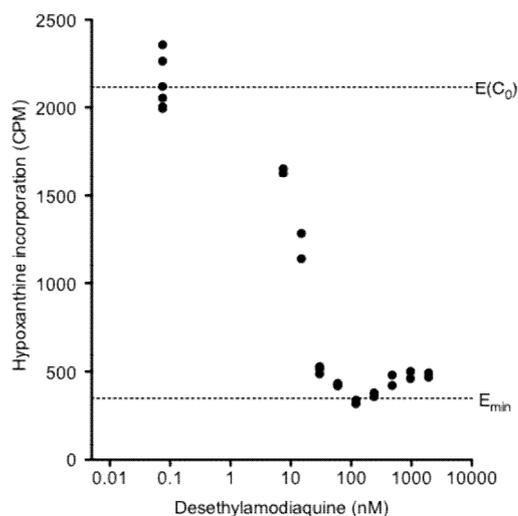


Figure 2: Example of primary (raw) data, indicating  $E(C_0)$  and  $E_{min}$ .

The y-axis shows hypoxanthine incorporation expressed as counts per minute (cpm). For display purposes, effect at zero drug is shown (at a concentration two logs below minimum concentration used). In this case  $EC_{max}$  (growth at the highest concentration) is higher than  $E_{min}$  indicating an apparent rise in growth at very high concentration.

### 5.2 Normalisation

Experimental data are normalized to a percentage scale based on the constraints  $E(C_0)$  and  $E_{min}$  (Figure 4).

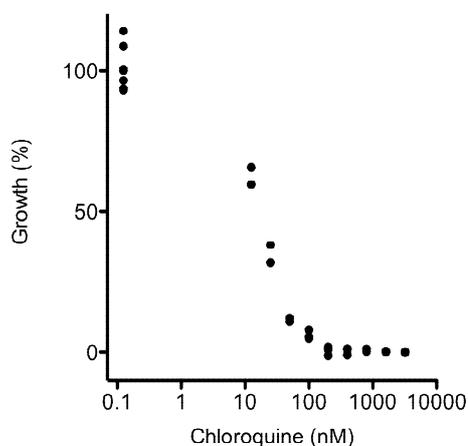


Figure 3: Normalisation of primary (raw) data to percentage scale based on  $E(C_0)$  and  $E_{min}$ .

For display purposes, effect at zero drug is shown (at a concentration two logs below minimum concentration used).

### 5.3 Non-linear regression

Non-linear regression using a sigmoid, 2-parameter, concentration-inhibition model is applied to the data based on established approaches (Le Nagard *et al* 2010), with the upper constraint set at 100% and the lower at 0% i.e.  $E(C_0)$  and  $E_{\min}$  (Figure 5). This final step yields two key results for each curve, the  $IC_{50}$  and gamma (an expression of the slope of the curve at the  $IC_{50}$ ), each with 95% confidence intervals.

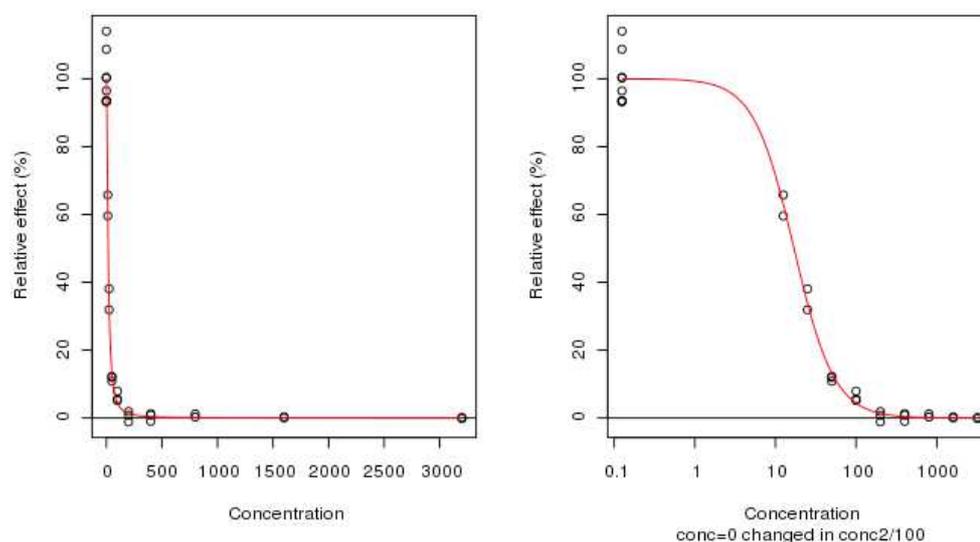


Figure 4: Non-linear regression, constrained to 100% and 0%, based on  $E(C_0)$  and  $E_{\min}$  respectively. The data are shown on both linear (left) and log (right) x-axis scales. For display purposes, the effect at zero drug is shown (at a concentration two logs below the minimum concentration used).

### 5.4 Core Criteria for summary analyses

Confidence in the  $IC_{50}$  value and slope resulting from linear regression will be used to define a subset of results with a tight confidence interval suitable for core analyses and reports according to the following core criteria:

- If gamma is not 10, ratio of upper : lower 95% confidence intervals for  $IC_{50}$  must be less than 3 (and both confidence intervals positive)
- If gamma is 10 (indicating a fixed, steep slope), ratio of  $E(C_0) : E_{\min}$  must be greater than 2 indicating acceptable growth (Basco 2007)

Note: On some occasions, non-linear regression fails to converge or produce an acceptable curve, due to either a very steep slope or noisy data. Under such circumstances, the non-linear regression algorithm repeats the process with a fixed gamma of 10. If this second iteration is successful, this renders the process a 1-parameter model. Such curves are consequently heavily constrained (at top, bottom and slope) so that the confidence intervals for  $IC_{50}$  (the last variable parameter) may be insensitive measures of true confidence. For this reason the additional criterion based on signal and noise is applied in such cases.

Available evidence indicates that these criteria are appropriate for many datasets, particularly those involving ELISA and SYBR green readouts. However additional parameters may be helpful for analysis of specific datasets. For this reason IVART also includes in the spreadsheet output the root mean square deviation (quantifying overall goodness of fit), a measure that

could potentially be applied in a systematic way by investigators under particular circumstances.

## 5.5 Range warnings

With some datasets there is evidence that the range of drugs used to determine IC<sub>50</sub> is too high or too low, and Range warnings are accordingly triggered. Range warnings are applied independently of core criteria so assays that meet core criteria, but trigger a range warning, are retained in core analyses.

The 'Range high' warning is triggered when  $IC_{50} < \text{lowest drug concentration}$ . The drug range is so high that even the lowest drug concentration already inhibits growth by more than 50%.

The 'Range low' warning is triggered when  $(EC_0 - EC_{max}) > (1.05 \times (EC_0 - E_{min}))$ . This is designed to pick up 'unfinished assays' where there is evidence that further inhibition would have occurred if a higher range of concentrations had been used.  $EC_0 - EC_{max}$  is the efficacy of the drug measured at the highest drug concentration while  $EC_0 - E_{min}$  is the normal measure of drug efficacy used by IVART. In assays where  $EC_0 - EC_{max}$  is more than  $EC_0 - E_{min}$  efficacy is continuing to increase at the highest concentration; a threshold of 1.05 corresponds to an additional increase in inhibition of 10% at  $C_{max}$  compared to any other concentration.

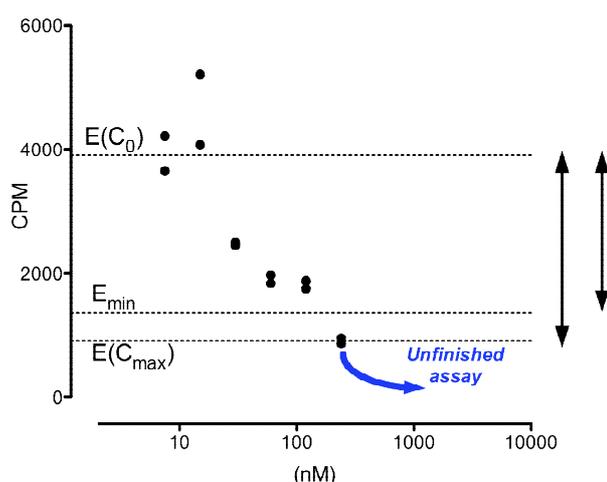


Figure 5: Measures of drug efficacy  
In this case  $(EC_0 - EC_{max}) > (1.05 \times (EC_0 - E_{min}))$ , indicating an unfinished assay. This would trigger the 'Range low' warning.

## 6. Reporting

### 6.1 Study report

An automated report in pdf format will be generated and returned to the contributor, describing transformed data and *in vitro* susceptibility results.

#### Summarised data

Tables will display the number of samples tested, assays undertaken and assays with successful non-linear regression, as well as the number of assays falling into the core subset with tighter confidence intervals (see 5.4). Results are presented according to the study date and drug used.

Box plots show median, quartiles and range for  $IC_{50}$  values for each drug and year (Figure 7). Only data where there are at least 10 results per drug per year will be plotted in this way.

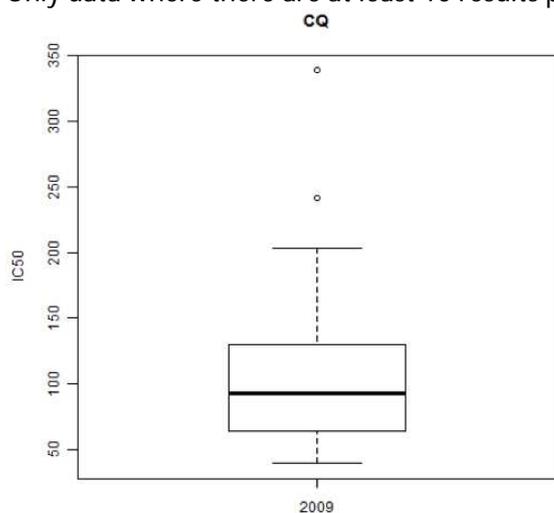


Figure 7: Box plot showing median, quartiles and range for  $IC_{50}$  values for chloroquine in 2009. Points more than one quartile width away from the main quartile are shown as outlying values.

#### Individual assay data

Each assay curve with individual normalized primary output data will be shown in the report to the contributor. For each sample, the  $IC_{50}$  estimate for each drug assay curve with its upper and lower 95% CI, the upper and lower constraints used, and gamma will be tabulated. Lists of assays where no regression analysis could be obtained, assays with no sample data and samples with no assay data will be provided.

## 6.2 Microsoft Excel™ files

Data contributors will receive spreadsheet files containing the output from IVART, including a single table summarising key parameters for all individual assays. It will integrate the original sample data supplied by the contributor allowing a variety of further analyses e.g. by location, drug, year etc..

## 6.3 Review by curator and contributor

The automated report is reviewed by the curator and then sent, as a draft, to the contributor for review and approval before locking as the final report.

## 7. WWARN Explorer

The [WWARN Explorer](#) is an online, open-access tool which allows users to perform custom queries of multiple studies and visualise the results using dynamic interactive maps. Following approval from the contributor, elements of the study report and box plots – shown by drug and year – will be displayed on WWARN Explorer.

## 8. Concluding remarks

This is the first version of the WWARN In Vitro DMSAP. Future versions will incorporate user feedback and knowledge gained as additional datasets are submitted to WWARN. A central goal of WWARN's In Vitro module is to link patient isolate data with results for validated

reference strain 3D7 obtained in the same laboratory. Such analyses will be incorporated into future versions of the In Vitro DMSAP.

## 9. References

- Le Nagard H, Vincent C, Mentre F, Le Bras J. Online analysis of in vitro resistance to antimalarial drugs through nonlinear regression. *Comput. Methods Programs Biomed.* 2010 Sept 7.
- Basco LK. Field application of in vitro assays for the sensitivity of human malaria parasites to antimalarial drugs. World Health Organization 2007.