# MAP INCIDENCE DATA GATHERING PROTOCOL

Author: Mike Thorn, Kate Twohig, Harry Gibson  
Owner: MAP Programme Manager

<table>
<thead>
<tr>
<th>Revision date</th>
<th>Author</th>
<th>Summary of Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-Jan-2019</td>
<td>MPT, HSG, KAT</td>
<td>First release</td>
</tr>
<tr>
<td>Chapter</td>
<td>Section</td>
<td>Table of Contents</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>------------------</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>1.1</td>
<td>Scope of document</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Motivation</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>SCOPE OF DATA COLLECTION / SEARCH</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>Geographic extent</td>
<td>3</td>
</tr>
<tr>
<td>2.2</td>
<td>Temporal extent</td>
<td>3</td>
</tr>
<tr>
<td>2.3</td>
<td>Temporal resolution</td>
<td>4</td>
</tr>
<tr>
<td>2.4</td>
<td>Types of data – API</td>
<td>4</td>
</tr>
<tr>
<td>2.4.1</td>
<td>First order of preference</td>
<td>4</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Second order of preference</td>
<td>5</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Third order of preference</td>
<td>5</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Fourth order of preference</td>
<td>5</td>
</tr>
<tr>
<td>2.4.5</td>
<td>Fifth order of preference</td>
<td>5</td>
</tr>
<tr>
<td>2.5</td>
<td>Considerations for representativeness</td>
<td>5</td>
</tr>
<tr>
<td>2.6</td>
<td>Types of data – other</td>
<td>5</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Mortality-related data</td>
<td>6</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Intervention data</td>
<td>6</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Health facility data</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>BIBLIOGRAPHIC / SOURCES INFORMATION</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>APPROACH TO DATA EXTRACTION</td>
<td>7</td>
</tr>
<tr>
<td>4.1</td>
<td>Special considerations</td>
<td>7</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Distinguishing confirmed and unconfirmed cases</td>
<td>7</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Calculating values</td>
<td>7</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Dealing with missing data and implied data</td>
<td>7</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Very high-resolution spatial data</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>SEARCH STRATEGY</td>
<td>8</td>
</tr>
</tbody>
</table>
GBD Incidence Data Gathering Protocol

1 Introduction

1.1 Scope of document

This document covers the protocols used by MAP for gathering data related to routine health surveillance activities and interventions. The primary type of data covered is “incidence data”; that is case data that can be used to calculate Annual Parasite Incidence (API) data or a species-specific subset (AFI, AVI). Where available, malaria-related mortality data and information on interventions is also sought.

These types of data are mainly available through national Ministries of Health (MoH), malaria control programmes and non-governmental organizations.

The document identifies the ways in which sources should be identified and recorded, and how data should be extracted.

Not all known sources of incidence data are covered in this protocol and it should be adapted to include more details of alternate data sources when those become known.

1.2 Motivation

The incidence data gathered by MAP are used by our modellers, in conjunction with our covariate datasets and parasite rate survey data, to model and predict malaria incidence globally within the administrative and demographic sub-divisions used in the GBD outputs.

2 Scope of Data Collection / Search

2.1 Geographic extent

For 2018 onwards, the scope of the data collection process is global. This contrasts with data gathering exercises conducted in previous years, where the data collection excluded most sub-Saharan African countries (a specific subset of these were included).

2.2 Temporal extent

Some forms of data are easier to extract than others; therefore, there is a trade-off between the effort required and the benefit it will bring to the modelling process.

For sources that are deemed “hard” to extract, such as PDFs (especially where these do not contain tables that can be easily copied), data are sought and extracted for a maximum period of ten years prior to the current data gathering round (i.e. 2007 – 2017 in 2018). However, any availability of data for a longer period will be noted.

For sources that are deemed “easier” to extract, such as online data portals or data published in spreadsheet form, there is no specific time limit.
2.3 Temporal resolution

As a minimum, annual data are required. Quarterly, monthly, or weekly data are preferred if available. Clearly extracting weekly data is much more work than extracting annual data, so decisions are made on a case-by-case basis of which data to extract, based on research priorities. Sources are in any case gathered at the best-available resolution, even if only a subset are initially extracted.

2.4 Types of data – API

As an output of the data gathering process, the minimum MAP requires for modelling, is calculated API and a proportion of Pf cases. Of these, API is derived, with varying levels of reliability, from several types of data that may be found in the data sources; whilst the proportion of Pf cases can, when necessary, be imputed from elsewhere, but this is less satisfactory than obtaining specific species breakdowns directly from the raw data itself.

The types of data whose existence is checked for in the sources have been divided into several tiers or orders of priority to maximise the reliability and utility of the data. Higher-tier data are collected in preference to lower-tier, although it is relatively unusual for there to actually be a choice of sources available for a given area / timespan.

2.4.1 First order of preference

- Data should be broken down at Admin 1 unit level as a minimum; preferably by Admin 2 or Admin 3
- As a minimum, a total number of confirmed and unconfirmed malaria cases should be given for each species (Pf, Pv, Other malaria); a breakdown by age and sex bands is preferable.
- Information on microscopy examinations should be given as follows:
  - Total number of microscopy examinations undertaken
  - Total number of microscopy examinations positive for Pf (broken down by age and sex bands if available)
  - Total number of microscopy examinations positive for Pv (broken down by age and sex bands if available)
  - Total number of microscopy examinations positive for mixed Pf and Pv (broken down by age and sex bands if available)
  - Total number of microscopy examinations positive for other malaria species (broken down by age and sex bands if available)
- Information on RDT examinations should be given as follows:
  - Total number of RDT examinations undertaken
  - Total number of RDT examinations positive for Pf (broken down by age and sex bands if available)
  - Total number of RDT examinations positive for Pv (broken down by age and sex bands if available)
  - Total number of RDT examinations positive for mixed Pf and Pv (broken down by age and sex bands if available)
  - Total number of RDT examinations positive for other malaria species (broken down by age and sex bands if available)
- An indication should be given of whether the microscopy and RDT tests relate to the same people
• A value for reporting completeness should be given
• A value should be given for the proportion of children under 5 with fever who sought treatment for the fever

2.4.2 Second order of preference
• Data should be broken down at Admin 1 unit level as a minimum; preferably by Admin 2 or even Admin 3
• As a minimum, a total number of confirmed and unconfirmed malaria cases should be given for each species (Pf, Pv, Other malaria); a breakdown by age and sex bands is preferable for each of these.

2.4.3 Third order of preference
• Data should be broken down at Admin 1 unit level as a minimum; preferably by Admin 2 or even Admin 3
• Pre-calculated values should be given for each of API, AVI, and AFI.

2.4.4 Fourth order of preference
• Data should be broken down at Admin 1 unit level as a minimum; preferably by Admin 2 or even Admin 3
• Pre-calculated values should be given for API and an indication of the percentage of these cases that were caused by Pf.

2.4.5 Fifth order of preference
• Data should be broken down at Admin 1 unit level as a minimum; preferably by Admin 2 or even Admin 3
• Pre-calculated values should be given for API.
• Note that in these cases we would use a national Pf case percentage figure to calculate AVI and AFI later

2.5 Considerations for representativeness

Consideration is given as to how representative the data is, including whether it provides complete geographic coverage of an entire administrative unit of interest (i.e. routine health information vs. reports from a subset of health facilities as part of a sentinel surveillance system).

Identifying completeness of the data in terms of coverage of a country’s administrative units is not as simple as identifying whether malaria data have been found for all that country’s known administrative units. For example some countries will only report on malaria for administrative units that have reported cases (implying zeros elsewhere), while others might be only reporting on a subset of their administrative units due to within-country political instability. As long as the administrative-level data appears to be complete, both of these scenarios would provide data that is useful and should be collected.

2.6 Types of data – other

Regardless of which order of API information is present according to the categorisation in section 2.4, the following other types of information are also collected if present:
2.6.1 Mortality-related data

- Number of deaths, with breakdowns by species, gender, or age-bands where available
- Case fatality rates
- Any other metrics relating to deaths

2.6.2 Intervention data

Any data that are present regarding antimalarial interventions will be extracted, according to the following points.

- **ITNs**
  - Number distributed by admin units at lowest resolution – national is acceptable
  - Amount spent on ITNs by admin units at lowest resolution – national is acceptable
  - Estimates of % coverage by admin units at lowest resolution – national is acceptable
- **IRS**
  - Quantity and type of insecticide used by admin units at lowest resolution – national is acceptable
  - Amount spent on insecticide by admin units at lowest resolution – national is acceptable
  - Number of households sprayed by admin units at lowest resolution – national is acceptable
  - Estimates of % coverage by admin units at lowest resolution – national is acceptable
- **ACTs**
  - Number distributed by admin units at lowest resolution – national is acceptable
  - Amount spent on ACTs by admin units at lowest resolution – national is acceptable
  - Estimates of % coverage by admin units at lowest resolution – national is acceptable

2.6.3 Health facility data

If any data are present regarding health facilities, these will be extracted according to the following points.

- The number of Health Centres, Posts, and Hospitals, split by public/private (if representative of the entire unit of interest)
- If available, the geographic coordinates of each facility

3 Bibliographic / sources information

Bibliographic and source information is recorded for all sources of data extracted and associated with those data in MAP’s databases.
4  Approach to data extraction

During the data extraction process, it is crucial that consideration is given to the relative quality of the data, rather than simply the quantity. In particular, it is likely that within any given report the authors will represent the same data from a number of different angles in order to illustrate the points being made by the text. It is not necessary to extract all these different representations of the data, but instead prioritize the lowest level disaggregated information, in accordance with the types of data laid out in section 2.4.

The list of data extraction considerations below is not exhaustive and any additional special situations should be added when encountered.

4.1 Special considerations

4.1.1 Distinguishing confirmed and unconfirmed cases

The definition of unconfirmed cases are those cases in a report listed as being suspected of being malaria but **no formal diagnostic test was undertaken to verify the presumed diagnosis**. However, the text used in the reports will seldom use the words “unconfirmed cases”. Sources should be carefully read to look for figures that can be interpreted as unconfirmed cases, such as “presumed cases” or “cases with fever given anti-malarial drugs”. Each source will have slightly different ways of describing these data. Also note that for many countries with more robust health systems, all suspected malaria cases are tested and so no unconfirmed case figures are reported (under any description).

The category “suspected cases” refers to those cases that were believed to be malaria **before** being tested, from which a subset of tests may have been performed. If no tests are performed, all “suspected cases” will be “unconfirmed cases”, however if all “suspected cases” are tested, then there will be no “unconfirmed cases”.

The definition of confirmed cases are those where a formal diagnosis has taken place using microscopy, RDTs, PCR, or any other method.

4.1.2 Calculating values

Whenever possible, MAP extracts exact values to allow for easier referencing of the original source. MAP’s databases are able to create aggregations once data is collected, so aggregations are not carried out during extraction.

4.1.3 Dealing with missing data and implied data

An exception to the above occurs when a data source is missing a subset of information that would make the database calculations inaccurate. In this case, the partial information should not be extracted as it might cause a misunderstanding and misrepresentation of the data.

4.1.4 Very high-resolution spatial data

Data that cannot be matched to MAP’s administrative shapefiles (i.e. lower than ADMIN3, or ADMIN2 in some cases) or does not have coordinates is not extracted.
5 Search Strategy

Routine surveillance systems data are usually published by ministries of health or NGOs in periodic reports. Such reports are not catalogued in traditional academic repositories such as PubMed. Sourcing such data therefore relies on internet searches to locate ministry and NGO web sites and to search these web sites for downloadable publications and figures.

A complication arises with web sites in languages not spoken by members of the MAP group, in particular sites that use non-Latin character sets. Fortunately, it is common for graphics for charts and graphs within reports to have their legends in western characters, regardless of the language used in the report.

MAP’s search strategy focuses around:

- Using a set of common terms for malaria in a variety of languages
- Using these terms to search for both text and images using web search engines

MAP has now established the locations of the ministry of health websites for most malaria-endemic countries. For each cycle of data collection, MAP:

- Revisits the established list of web sites to extract any newly released reports
- Undertakes a search for any new web sites