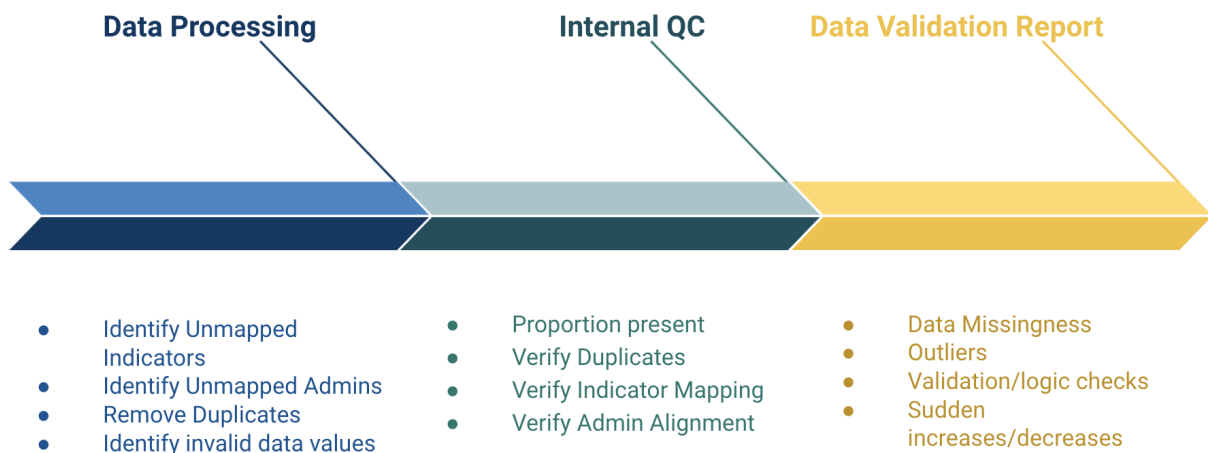# Malaria Quarterly Report Quality Control Processes Overview

Updated July 2022

The PMI [Malaria Quarterly Report](#) (QR) is the primary data resource in M-DIVE for understanding malaria incidence, burden, and progress metrics in PMI-supported countries. It includes data provided by countries each quarter, based on what they have available from their existing data collection systems. This document includes an overview of all of the quality control checks that are conducted on the QR data once they have been submitted to the Civs Analytics M-DIVE team.

The QR data are submitted, processed, and standardized into M-DIVE through our QR Data Pipeline to ensure consistency of format, time intervals, geographies, and indicators. (For more, see [this overview of the QR Data Pipeline](#).) Quality Control (QC) occurs in two main phases: **during data processing** and **post data processing**.

Data Processing
- Identify Unmapped Indicators
- Identify Unmapped Admins
- Remove Duplicates
- Identify invalid data values

Internal QC
- Proportion present
- Verify Duplicates
- Verify Indicator Mapping
- Verify Admin Alignment

Data Validation Report
- Data Missingness
- Outliers
- Validation/logic checks
- Sudden increases/decreases

Data Processing QC Checks

During the data processing phase, the Civis team identifies and rectifies known data quality concerns related to the submitted source data, through both manual inspection and automated tools built into the QR Data Pipeline.

1. **Identify Unmapped Indicators** - Countries may use different names for the same indicator (e.g., text-difference, different language) or have it broken down in different ways (e.g. 'Total Malaria Deaths' vs 'Malaria deaths among <5' + 'Malaria deaths among >= 5'). To be able to compare indicators across different countries (and quarters), source data indicators are mapped to a shared, standard set of names and/or combine indicators where

necessary. This QC check inspects the source data and provides a list of indicators in the submitted data that have not been previously mapped to a standardized M-DIVE indicator, so that the Civis team can assign an appropriate standardization before the QR pipeline is run.

2. **Identify Unmapped Administrative Subdivisions**- Geographical area data are sometimes submitted with considerable variation (e.g. differences in spelling, language) and sometimes administrative boundaries change from data collection round to round. This QC check inspects the source data and compares the reported geographical areas to a list of standard geographical areas provided to Civis by the in-country teams to look for discrepancies (ex. spelling or new geographical boundaries that are not yet captured in the standard list). The QC check then provides a list of reported geographical areas that need to be mapped to standard geographical areas.

3. **Remove Duplicates** - Duplicates are either present in the source data, or they get introduced during the automated processing phase due to geography alignment errors (e.g., two different areas that get mapped to the same standardized area) or time standardization (e.g., data from a weekly source that, when aggregated to the monthly level, becomes a duplicate of data from a monthly source). This check removes the duplicate data.

4. **Identify invalid data values** - This check is done as part of the automated portion of the data pipeline, and it highlights invalid data points such as out-of-sample values (e.g., month = 15, year = 1930) and drops them before appending the data to the final reporting table. When an indicator is submitted entirely with blank rows, the indicator is interpreted as not submitted and filled in as nulls in the reporting table for the country and reporting period in question. When an indicator is submitted partially with some blank rows, the non-null values are retained and the blank rows are filled in as nulls in the final reporting table. If inspections indicate that a country submitted 0s as placeholders for nulls, the Civis M-DIVE team reaches out to the in-country team to confirm and these 0s are then converted to nulls. In some cases, an indicator is submitted with "N/A" indicating a special meaning such as the country doesn't participate in the program, doesn't use a particular drug, or doesn't collect information on a particular indicator. At this time the pipeline processes these "N/A's" as nulls, but there are plans to eventually differentiate between "N/A" due to special circumstances and nulls due to blank submissions.

Post Processing QC Checks

Once data have been processed into M-DIVE, the Civis M-DIVE team conducts further quality control checks to identify additional issues that may arise in the processed data, before releasing the final data that is then surfaced in reports and dashboards. Some common issues may be resolved directly by the Civis team, but most types of data quality concerns or questions will be shared with the corresponding PMI country team backstops to ensure that data are being reported and interpreted correctly.

CIVIS®

Quality Checks Performed by Civis on the Processed Data:

1. **Proportion Present** - This check highlights indicators that were submitted in the source data, but did not make it into the processed table for all expected geographies and time periods. Issues identified in this step are usually addressed by the Civis team correcting or validating that submitted indicators are correctly mapped to the standard values expected during the processing round. This check also shows the percentage of data present in the processed table for each indicator for the round of data submission, from which the Civis team can investigate why certain indicators have lower coverage than expected.

2. **Verify Duplicates** - This check highlights the duplicate records that were dropped during the automated processing phase for approval/confirmation by an analyst.

3. **Verify Indicator Mapping and Geography Alignment** - We perform a second set of checks for indicator mapping and geography alignment on the processed table to make sure that no source data indicator went unmapped to a standard indicator or geographical area during the data processing phase.

Data Validation Report:

The data validation report is available to all users to inspect and do quality control checks of the final data that is surfaced in M-DIVE.

1. **Data Missingness** - This check shows what data are and are not available in the final processed QR data table in M-DIVE. It does not distinguish data that is intentionally omitted or not applicable, such as indicators that might not be relevant to a given country, but it can be used to deduce if there are trends in missingness in indicators or geographical areas across various time periods.

2. **Outliers** - For each country and indicator, we highlight any specific data points that are significantly outside of the normal or expected range[1].

3. **Validation/logic checks** - A series of logic checks are conducted based on assumptions about the data and malaria epidemiology. For example, we check to see if the number of tested cases is less than suspected cases, the sum of reported male and female population is equal to the reported total population in a geographical area, etc.

4. **Sudden increases/decreases** - This check highlights indicators that have significant increases or decreases compared to a previous time point (e.x. previous month, same season in the previous year, etc.).

External Validation

Once our internal processes are concluded and data are approved for inclusion in M-DIVE, the Data Validation Report is made available to PMI country users who are encouraged to discuss and

---

[1]Normal range is defined as any observation within 10 times the Interquartile Range (IQR). The IQR is a statistical measure of dispersion and is calculated using all the data available for a country.

investigate records highlighted as invalid as part of national data validation processes or during monthly or quarterly data review meetings with the National Malaria Control Programme and/or Ministry of Health.

If the data quality issue highlighted in M-DIVE can also be found in the source data, and the country decides to make changes to their dataset, the country should re-submit data for the entire period in which updates were made.

If you have any questions, please submit them to Alaine Knipes (aknipes@cdc.gov) and Misun Choi (mchoi@usaid.gov) on the PMI Surveillance & Informatics team, or to support@civisanalytics.com. If you would like to learn more about additional topics on M-DIVE, please click the links below for further reading.

Important Resources

- [Malaria Quarterly Report Data Pipeline Overview: How Quarterly Report Data Are Processed in M-DIVE](#)
- [Malaria Quarterly Report Time Standardization](#)
- [Malaria Quarterly Report Indicator Standardization](#)
- [Malaria Quarterly Report Format Standardization](#)
- [Malaria Quarterly Report Geographic Standardization](#)
- [Malaria Quarterly Report](#) (M-DIVE access required)
- [M-DIVE Help Center](#) (M-DIVE access required)