For aim 2a we, shall follow the definitions proposed by Kahn and operationalize them to this study. The data quality dimensions are assessed based on validation and verification. Verification has the key feature to determine expected values and distributions using resources within a local environment. On the other hand, validation focuses on alignment of data values with respect to external benchmarks. We present the dimension definitions in these two contexts below:

**Completeness:** This will be measured by assessing the frequencies of data attributes in a data set without reference to the data values. The restriction to assessing only the presence or absence of values without assessing whether these values fall in allowable limits differentiates completeness to plausibility. Completeness may be measured as a single instance or over several points in time. These time variations of completeness will be applied differently based on the data variables to be assessed in this study.

**Conformance**: This focuses on the agreement of the data against internal or external formatting, relational or computational definitions. This characteristic does not attempt to measure the completeness or plausibility of the values observed. In this case, we shall assess selected data variables for their conformance to definitions as specified in a data dictionary for the EHR database in Malawi. Conformance may be assessed as value conformance: assesses if recorded value meets a predefined data architecture definition. These could be data domains, for example in terms of patient address, only having data from specific areas. Relational conformance checks whether data elements meet constraints imposed by database definitions, for example ensuring that relational integrity requirement is met. An example could be assessing that internal patient ID for an encounter table has a corresponding ID in a drugs dispensation table. Computational conformance on the other hand assesses whether calculated values are correct or match the expected definitions.

**Plausibility:** This deals with believability or truthfulness of values of data. Unlike completeness and conformance that deal with presence and structure of data values, plausibility deals with how closely data values represent the world.

Plausibility therefore may be assessed as unique plausibility: This deals with finding duplicate values of a unique value representing two real world things, atemporal plausibility: this aims to seek if observed data values meet with internal expected values such as range checks or externally validated density distributions, such as weight for height checks for example. Temporal plausibility assesses whether the observed values follow expected progression, with an example of weight for age values as an example. Another example would be precedence of changing regimen and observed or documented reason for changing that regimen. The specific definitions for the analyses are in table 10 below and the level of plausibility shall be confirmed with national guidelines for treating HIV/AIDS(88).

In the analysis of the characteristics above, descriptive statistics such as proportions, count statistics and charts will be used for the data quality assessment. The specific definitions for the analyses are in table 12 below. The definitions below are mainly focused on verification unless specified in the table. This is due to the EHR being the main primary data source.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **90-90-90 data use** | **Variable** | **Completeness** | **Atemporal Plausibility** | **Value Conformance** | **Completeness (Validation)** |
| Identification of HIV positive individuals | Unique patient identifier | all assigned patient identifiers/all registered patients with a dispensation | All patient identifiers not compliant with format/all patient assigned identifiers | All patient identifiers issued at registration with matching data dictionary definition/all patient identifiers in the database | All patient identifiers issued at registration verified at central database/ All patient identifiers issued for that site |
| Patient tracing | Patient Address – current place of residence  Patients with contact phone number | All patients registered with a current place of residence/all patients registered with a dispensation  All patients registered with a phone number/all patients registered with a dispensation | All non-existent current places of residence /all patients with a current place of residence  All patients registered with a wrong sequence of a phone number/ All patients registered with a phone number | All patients with current place of residence matching domain of residences documented at registration/all patients registered with a dispensation  All patients registered with a phone number documented at registration with correct database format/ all patients registered with a dispensation | All patients with current place of residence documented at registration verified with master/ All patients with current place of residence documented at registration |
| Quality of patient care | Prescription versus dispensed drugs  Patients with changed regimens and reasons for changing | Number of prescriptions filled/number of prescriptions made  Number of patients with changed regimen and documented reason for changing/Number of patients with changed regimen | Number of prescriptions matching dispensations/number of dispensations made  Number of patients with documented reason for changing that does not match the expected recommendation/ number of patients changed to a specific regimen and documented reason for changing | Number of prescriptions filled on visit with correct database definition/Number of prescriptions filled  Number of patients with changed regimen and documented reason at changing saved from domain of reasons/number of patients with changed regimen and documented reason for changing | N/A  N/A |
|  | Body mass index – weight and height | Number of patients without height/number of patients registered with a dispensation | Number of patients with weight that is out of bounds/number of patients with weight and height  Number of patients with height that is out of bounds/number of patients with weight and height | Number of patients with weight and height documented at registration with expected domain range/number of patients with weight and height documented | N/A |
| Viral load suppression | Viral load results present | Number of patients with one observation of viral load result 6 months or more after registration/all patients with a registration into the ART program 6 months or more after registration | Number of patients with out of range viral load results/Number of patients with one observation of viral load 6 months or more after registration | Number of patients with a viral load result within expected timeframe of milestone with correct domain range/all patients that reached a viral load milestone within 6 months or more after registration | Number of patients with viral load that match with Central Database/Number of patients with viral results for a site 6 months or more after registration |

Table 13: Definition of Data Analysis Proportions

To ensure validity of the observed data quality measures in question 2a, we shall present the data quality measures to selected key informants. This will be done to ensure that the measures are correct.