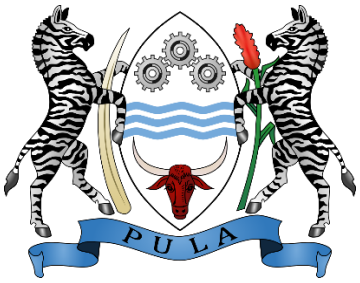
Standard Operating Procedures

Data Management Plan for the Botswana HIV Case Based Surveillance Protocol



Republic of Botswana

MINISTRY OF HEALTH & WELLNESS

Written by: Case Base Surveillance (CBS) Technical Working Group (TWG)	
Reviewed by: Akeem Ketlogetswe and Chris Serumola	**Current Version:** 1.0
Approved by: Penny Makuruetsa	**Effective Date:** 01/07/2021

Date of Development: July 2021
Updated: October 2021

1.0 TITLE

Data Management Plan for the Botswana HIV Case Based Surveillance Protocol

2.0 PURPOSE

- To define procedures for data monitoring and management of data collected during the implementation of the HIV Case Based Surveillance (CBS) protocol in Botswana.

- To define the processes involved in reviewing, monitoring, and improving the quality of data collected for HIV CBS in Botswana.

- To describe the procedures for communicating with health facilities on any data queries as well as providing feedback for improvement of HIV services.

3.0 INTENDED USERS

3.1 This standard operating procedure applies to all who have direct involvement in data monitoring, during the lifetime of the CBS surveillance activity. This will include health facility staff (MOHW and implementing partners (IP)), national data warehouse (NDW) officers, monitoring & evaluation (M&E) officers, district focal persons, trainer of trainers and monitors, CBS master trainers, and CBS technical working group (TWG) members.

4.0 RESPONSIBILITIES

4.1 Health Facility Staff

All health facility staff who play a role in caring for persons living with HIV (PLHIV) at the facility level are required to support and adhere to the execution of the procedures delineated in this document. The following field staff must make every effort to document clinical events either in the electronic medical record (EMR) or on the CBS case report forms (CRF) and store all client records securely and confidentially:

4.1.1 *HIV Testing Services Healthcare Assistants (HTS HCA)*

Provide HIV testing services (HTS), and collect results of the HIV test. Results must be documented in either: (1) paper CRF or (2) directly into DHIS-2 instance on tablet/computer or (3) in the EMR used at the health facility.

An HIV test resulting in an HIV positive diagnosis triggers a sequence of HIV care related activities. HIV diagnosis is the first sentinel event and the entry point into HIV case surveillance; this must be correctly documented in the tools identified above. Healthcare assistants (HCA) will also be involved in the data quality improvement (DQI) processes.

4.1.2 *Data Clerks/Data Support Officers (DSOs)*

Maintain adequate and accurate participant records (electronic and paper copies) containing all information collected during a clinical or laboratory visit for each PLHIV. Ensure all paper CRF in the facility are validated and then captured into DHIS-2 in a timely manner. Ensure that all CRF are filed and available for the monitors to inspect. Responsible for data cleaning processes at the facility level for all data issues identified at the NDW.

4.1.3 *Clinical Staff*

Maintain adequate and accurate participant records (electronic and paper copies) containing all information collected during a clinical or laboratory visit for each PLHIV. For facilities with an active EMR, the clinicians must ensure full system utilization in capturing correct patient data in a timely manner. During system downtime the clinician will capture patient data in paper-based system then later transition the information into the system when its active again. For paper-based facilities they will ensure all paper CRF in the facility are packaged weekly and transported to mother facility or DHMT office for entry into the DHIS-2 according to the chain of custody SOPs. Ensure that all documents are available for the monitors to inspect.

4.2 District Level Staff

All District level staff who play a role in the implementation of the HIV CBS protocol are required to adhere to and support the execution of the procedures outlined in this document. The

following district level staff must make every effort to store all study records securely and confidentially:

4.2.1 *District Trainer of Trainers (TOT)*

Responsible for the implementation of HIV CBS in the district. Train facility staff on the HIV CBS protocol and documents. Develop a district specific implementation plan and monitor its implementation. Conduct district data review sessions and participate in data quality improvement processes for CBS data from the district. Train facility staff on EMR use and provide support and mentoring on the systems used. Ensure availability of CBS data at district level.

4.2.2 *District M&E Officers*

Provide supervision to DSOs and data clerks and serve as the first stratum of oversight to ensure adherence to procedures that are delineated in this document. Main contact point for CBS data at the district. Responsible for validating data captured on the electronic system by data clerks; the validation will be done by comparing the monthly DHMT report with the CBS dataset report. Provide feedback to and work with the data clerks on resolving any data issues identified by the analysis team. Share CBS reports with District Health Management Team (DHMT).

4.2.3 *District IT Officers*

Provide and maintain system hardware for facilities. Ensure connectivity and functionality of the system. They are the contact person for EMR functionality at the district level.

4.2.4 *Implementing Partners (IP)*

Maintain adequate and accurate participant records (electronic and paper copies) containing all information collected during patient visits for each PLHIV. For facilities with an active EMR, the clinicians must ensure full system utilization in capturing correct patient data in a timely manner. During system downtime the clinician will capture patient data in paper-based system then later transition the information into the system when its active again. Ensure that all documents are available for the monitors to inspect.

4.2.5 *DHMT Coordinators*

DHMT coordinator will oversee the implementation of CBS in their districts. They will receive reports and be involved in the district data review sessions to ensure that the findings and actions to improve HIV services provision in the district. CBS reports will be shared with the DHMT coordinators for their review and action delegation.

4.3 National Level Staff

MOHW Program officers and technical partners involved in the implementation of the HIV CBS protocol will have the following roles to play in the implementation of the HIV CBS protocol:

4.3.1 *CBS Master Trainers*

4.3.1.1 Responsible for training district TOTs on the HIV CBS protocol and documents.

4.3.1.2 Collaborate with the district TOTs to develop the district specific implementation plan and monitor its implementation.

4.3.1.3 Conduct district data review sessions and participate in data quality improvement processes.

4.3.2 *National Datawarehouse Administrators and Developers*

This team consists of database administrators and software developers. They are responsible for ensuring seamless data transmission from EMRs into the NDW.

4.3.2.1 Ensure extraction of required data elements from NDW into the CBS dataset for analysis by the data analysis team.

4.3.2.2 Crossmatch and de-duplicate submitted cases to ensure that unique cases are only counted once.

4.3.2.3 Develop DHIS-2 instance for data capture for sites without EMRs and provide maintenance support.

4.3.2.4 Implement dashboards for data reviews at site and national level. Inconsistencies and incomplete data will be communicated back to the NDW for verification.

4.3.3 *National M&E Officers*

Consists of M&E officers from both MOHW and IPs who:

1. Prepare the CBS dataset for analysis.
2. Provide feedback and collaborate with the NDW administrators and district M&E officers to resolve any data issues identified by the analysis team.
3. Ensure that data is being pushed into CBS database as expected, report any issues to the NDW, and assist with the troubleshooting.
4. Develop and maintain the data dictionary.
5. Collaborate with the analysis teams to develop products for dissemination including summary reports and the national epidemiology profile.
6. Review CBS data on a monthly basis for completeness, produce frequency tables to show completeness of data, and review the validity of data.
7. Provide feedback to sites on data quality gaps and ensure the timeliness of data submissions, representativeness of case data (are all health facilities submitting data) and prioritize follow-up activities for health facilities.

4.3.4 *Health Informatics Staff*

Create dashboards for data reviews at national level. Manage health information systems and monitor system utilization across all districts.

4.3.5 *Data Analysis Team*

This team consists of the MOHW Surveillance Officers and IP data analysts, who provide support in the form of data analysts, biostatisticians, and epidemiologists. This team will develop the CBS data analysis plan, provide guidance to the NDW team on both data extraction and the development of the dashboards. This team will prepare and submit quarterly and annual CBS reports to the CBS TWG and identified stakeholders. They will be using advanced statistical methods to account for missing data and/or site-level correlation whenever necessary. This information will be reported as part of the routine report, including with limitations of the data.

4.3.6 CBS Technical Work Group (TWG)

CBS TWG is comprised of MOHW and IP program officers, who will review and sign off on documents and reports prepared by the data analysis team. The team that conducts the data monitoring consists, but is not limited, to the following members:

a) **National M&E Officer:** Responsible for reviewing CBS data on a monthly basis for completeness, produce frequency tables to check for completeness of data and validity of data. Provide feedback to sites on data quality gaps. Ensure the timeliness of data submissions. Evaluation representativeness of case data and prioritize follow-up activities for health facilities. Preparation of the CBS dataset for analysis. Provide feedback to and working with the NDW administrators and district M&E officers to resolve any data issues identified by the analysis team.
b) **Data Analysis Team:** This team consists of the MOHW surveillance officer with IP data analysis support in the form of data analysts, biostatisticians, and epidemiologists who are responsible for monitoring analysis to summarize. Prepare and submit quarterly and annual CBS reports to the TWG and identified stakeholders.
c) **District M&E Officers:** Resolving all issues identified by the National M&E officers while working closely with the facility staff.
d) **District ToTs:** Conduct district data review sessions and participate in data quality improvement (DQI) processes for CBS data from the district. Train facility staff on EMR use and provide support and mentoring on the systems used.
e) **Facility Staff:** Includes clinicians, HCA, and data clerks responsible for resolving all issues identified by the district M&E officers.
f) **CBS TWG:** Review the data quality indicators produced by M&E officers as well as on the dashboard. Monitor data quality of CBS data.

5.0 PROCEDURES

5.1 Data Quality Monitoring

Data quality is the ability of data to serve its intended purpose, in each context within a specified process. It is based on the following distinct dimensions:

- **Accuracy**: Is the data correct, precise, and error-free? Without accuracy, data is misleading and useless.
- **Availability**: Is the right data available to the correct people within the organization? Data must be available and accessible for the people who need it so they can conduct their required activities.
- **Completeness:** Is the data complete? Is some information missing? Incomplete data leads to gaps in information, making it harder to effectively put data to use.
- **Granularity**: What's the level of detail that the data can provide? The right degree of granularity in data is necessary for accurate and effective decision-making.
- **Relevance**: Is the information collected needed? What's the purpose of the data stored? Irrelevant information just ends up wasting time, effort, and money.
- **Reliability**: Is the data ambiguous, vague, or containing any contradicting information? In all such cases, the information is unreliable, and the data cannot be trusted.
- **Timeliness**: Is the data outdated or obsolete? Data collected at the right time is an important measure of data quality. Relying on data that isn't timely is misleading and can lead to inaccurate decisions.

Data Quality Monitoring Process

1. The National M&E officers guides NDW administrators and developers to extract the CBS dataset.
2. The National M&E officers reviews the CBS dataset to determine if it gives the intended results.
3. The National M&E officers run a deduplication process to ensure that the CBS dataset is unique, using Excel or other applications that are available.
    - When running the deduplication process for the CBS dataset, the fuzzy logic method may be used (such as ExisEcho, an excel-add in). Create an identifier column by concatenation (name, surname, sex, DOB) that you will use to check for possible duplicates. Then run the data set on fuzzy logic set to 85% proximity to extract all the possible duplicates.

4. The National M&E officers communicate and share the possible duplicates with NDW administrators so the possible duplicates can be merged on the CBS extraction. Communicate with district M&E officers for data merging and cleaning.
5. The National M&E officers identify data quality issues from the CBS dataset looking at variables such as patient demographics, testing, initiation data, viral load, refills, clinical visits, and patient care status.

Identifying Data Issues:

1. Identify and highlight all the missing data from the CBS dataset.
2. Identifiers:
    - **For citizens**; Omang Number – check if the ID conforms to the standard requirement i.e. 9 digits with middle number (5th number) being either 1 or 2. The 1 is for male and 2 is for female for citizens.
        i. Using excel to check length of the ID with formula, (=Len) and (=mid) for middle digit.
    - **For non-citizen** only highlight clients with missing identifiers.
3. Test data and initiation data
    - Identify and highlight test dates before the year 1985, initiation dates before the year 1999 (to confirm) and future dates (i.e., dates that are beyond the current date during data cleaning).
    - Highlight records that do not meet the below criteria using an IF statement on Excel or any other method:
        - Compare the test date to DOB to ensure test date is not earlier than the DOB.
        - Compare the initiation date to DOB to ensure initiation date is not earlier than the DOB.
        - Compare the test date to initiation date to ensure initiation date is not earlier than test date.
4. Viral Loads
    - Identify and highlight invalid VLs for active clients during the reporting quarter; validity requirement for VL is 6 months.
    - Identify incomplete VL (i.e., not done, insufficient, or result pending – and immediately notify concerned facility.

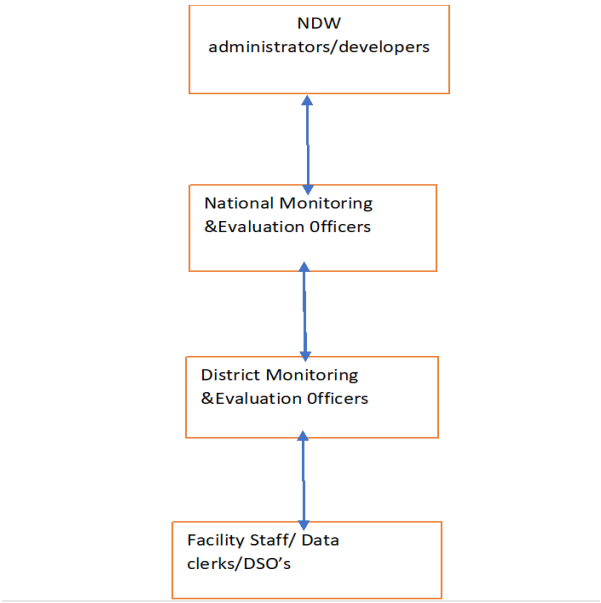
- Identify detectable VLs for active clients and notify facilities for management of detectable clients at the facility.

5. Refill and Clinical visits
    - Identify and highlight active clients with no refill dates, highlight clients with missing/incomplete regimens, and check for active clients with no latest encounter data recorded on EMR. Notify facilities to update the patient record(s).
6. Patient care status
    - Identify and highlight clients with no care status. Notify the facilities to update the client's correct care status accordingly in EMR systems (e.g., On ART, death, lost to follow up, defaulter, transfer out).

Sharing Identified data Issues with health facilities:

- The National M&E officers share the CBS data issues with the district M&E officer. Issues are separated into active and inactive clients for facilities to prioritize active client's issues. Instructions are clearly defined (e.g., issues are highlighted in yellow color, so the DSOs are to highlight in green when they have fixed the issue on the EMR. Highlight in blue when the issue is already up to date on the system). This is to help improve data extraction from NDW.
- The District M&E officers then share the identified CBS data issues with the facilities and data clerks to attend.
- Give feedback to the NDW administrators, where the issues identified are not seen as issues at the EMR system or if the dataset is inconsistent.

Data Flow



Feedback:

- Develop a feedback tool or use one that is currently in use to communicate the achievements, challenges, and way forward on identified data issues at all levels. (**Appendix 3**: Data Issues Feedback Reporting Tool).
- The CBS data manager compares the previous CBS report with the latest CBS report on progress made using data quality monitoring reports and dashboards.
- National M&E officers compiles and shares progress report from districts on identified data issues and report to the management on weekly basis.

Support and Mentoring on Data Quality:

- The National M&E officers follows up to monitor the data cleaning process with district M&E officers.
- The National M&E officers will do planned and unannounced site visits to districts. Districts with high data issues will be prioritized to ensure capacity building and transfer of skills.
- The National M&E officers will conduct virtual meetings, and attend to issues facilities face when resolving data quality issues and sharing best practices.

5.2 Data Quality Improvement

1. The data quality improvement (DQI) process entails a cluster of decisions centered on organizational data quality goals that determine the (1) data processes to improve, (2) solutions to implement, and (3) people to engage.
2. The best practice is to formalize the data quality improvement process. This can be done by properly documenting and tracking all process activities. This allows for progress tracking throughout the entire process.
3. The following process steps improve data quality:
    - **Establish a data quality environment**; refer to the dimensions of data quality.
    - **Assess data definitions**; clear definitions of the existing metadata, you cannot tell if something is wrong unless you can define what is right.
    - **Collect quality facts**; data profiling to discover the true content, structure, and quality of data. Identify record data related issues.
    - **Assess impact**; development of defining and deploying corrective actions after identifying data issues
    - **Identify causes**; before remedies can be produced, there is a need to identify the root causes of the identified wrong values. Only then can improvements in quality really structurally occur.
    - **Propose and implement remedies**; remedies are changes to systems that were previously designed and built in order to (1) prevent data inaccuracies from happening (again) in the future and (2) detect as many inaccuracies as possible when they occur.
    - **Monitor the results**; monitoring of the systems after remedies have been implemented for the positive impact and any occurrence of new problems.

Data Quality Improvement Process

The following steps describe how the DQI will be carried out to ensure quality data is collected from the manual patient files to the EMR systems and eventually to the NDW. The DQI will be done at both facility level and national level, by the data clerks and M&E officers, respectively.

Facility level

1. The data clerk officer will randomly select active patient files from the cabinet, a minimum of 10 to 50 files depending on the size of the facility.
2. The data clerk will check selected variables for consistency on both the patient file and the EMR system.
3. These variables may include but not limited to patient identifier, demographics, test date, initiation date, recent CD4, recent VL, last clinical date, and next appointment date.
4. The data clerk will tally the results showing consistent and inconsistent recordings, and also identify any possible causes for remediation.
5. All the variables checked for consistency between the patient file and the EMR system should be similar, or the EMR should have more up to date documentation.
6. The data clerk also calculates the selected indicators to check for consistency e.g., initiations for the specific period. The data clerk will tally if the newly initiated clients are correctly captured as such on the EMR system and record any data quality issues identified.

National level

1. The M&E officer will randomly select active patient files from the EMR system, a minimum of 10 to 50 files depending on the size of the facility.
2. The M&E officer will check selected variables for consistency on both the EMR system and the CBS dataset from the NDW.
3. These variables may include but not limited to patient identifier, demographics, test date, initiation date, recent CD4, recent VL, last clinical date, and next appointment date.
4. The M&E officer will tally the results showing consistent and inconsistent recordings, and also identify any possible causes for remediation.
5. All the variables checked for consistency between the EMR system and CBS dataset should be similar, or the CBS dataset should have more up to date variables.
6. The M&E officers will calculate selected indicators from the EMR system and the CBS dataset to compare for data consistency and document the findings. (See Appendices for proposed DQI tool).

Appendix 1: Monitoring Process

Criteria	Process	Level and Frequency	Result
Completeness	This requires: review of submitted case data in the surveillance database	Facility level staff may review their own data for completeness prior to submission (weekly/monthly) National level staff should review all newly submitted case data in real time, and the entire database periodically (monthly/quarterly)	Results should be presented by the service provider and system as a whole, describing the proportion of case data submissions with key data elements complete
Timeliness	This requires: comparison of two steps in the case surveillance process	Service provider staff may review their timeliness by looking at the sentinel events at the time of HIV diagnosis (confirmed HIV infection status, 1st CD4, OI assessments, and ART initiation) vs. the date of submission of the case data (monthly) District staff should review the time between key process events (quarterly) such as: Receipt of case data and data entry	Results should be presented by service provider and system as a whole, describing how their case data submission timeliness compares to the national standard
Accuracy	This requires: validation of data in the surveillance system against source data	Service provider staff may assess their own accuracy (monthly/quarterly) by comparing data submissions against the register National level staff should validate a sample of submitted case data against facility-level data periodically (quarterly/annually) with supervision visits	Results should be presented by service provider and system as a whole, describing common errors and places for improvement
Validity	This requires: review of data elements in the surveillance database	Service provider staff may review their own case surveillance data for validity before submission (weekly/monthly) National level staff should review all case surveillance data submissions (daily/weekly) and the database periodically (monthly/quarterly) for invalid values	Results should be presented by service provider and system as a whole, describing common errors and places for improvement
Representative-ness	This requires: comparison of cases in surveillance system against cases in source data (e.g., site registers)	Service provider staff may assess their own representativeness by comparing the number of diagnosed PLHIV whose case data were submitted to the system vs. number of diagnosed PLHIV in site records. National level staff should asses case surveillance representativeness (quarterly) by taking a sample of case data (e.g. diagnosis, ART initiation) at a service provider facility and verifying that it has been submitted to (is present in) the surveillance system.	Results should be presented by service provider and system as a whole, describing gaps in the submission of case data and points for improvement

Appendix 2: Monitoring Plan Outline

System Level	Monitoring Activities	Who	When
National and/or sub-National Level: Using: The master surveillance database Summary monitoring data generated at the national level (or sub-national, depending on data access)	Cross-match and de-duplicate submitted cases to ensure that unique cases are only counted once	National M&E officer	Weekly
	Produce frequency tables to check for: Completeness of data Validity of data Provide feedback to sites on data quality gaps	National M&E officer	Weekly
	Monitoring analysis to summarize Timeliness of data submissions Representativeness of case data (are all service providers submitting data?) Prioritize follow-up activities for service provider sites	National M&E officer and Surveillance Coordinator	Quarterly/ yearly
	Disseminate monitoring summaries to sub-national surveillance staff and service providers (quarterly) and stakeholders (yearly).	National M&E officer and Surveillance Coordinator	Quarterly/ yearly
Service Provider	Site monitoring visit to collect data on: Accuracy of submitted data Representativeness of submitted data (is the service provider submitting all sentinel events on all diagnosed PLHIV?)	Surveillance Coordinator or DHMT coordinators or National M&E officer	Quarterly/ yearly
	Direct observation of case data submission processes: Adherence data submission procedures Adherence to privacy, security and confidentiality policies Clearly defined data submission responsibilities and timelines Supervision of case surveillance activities Barriers to optimal system functioning	Surveillance Coordinator or DHMT coordinators or National M&E officer	Quarterly/ yearly/ as needed
	Follow up observation, training or support to investigate or remediate data quality or monitoring issues.	Surveillance Coordinator or DHMT coordinators or National M&E officer	As needed
	Routine monitoring of: Case data submission processes Quality of submitted case data (completeness, validity, representativeness, accuracy). Adherence to privacy, security and confidentiality policies	Facility Heads	Monthly

Appendix 3: Feedback Reporting tool

Data Issues Feedback Reporting Tool

District:_____ **Facility**: _____
Reporting Period:_____ **Reporting Officer**: _____

Data Issues -Variable	**Data issues - Baseline**	**Updated/fixed data issues on EMR**	**Challenges**	**Additional Comments**
Patient identifiers • *Patient demographics*				
Testing data • *Test date* • *Facility tested*				
Initiation data • *Initiation date* • *Facility initiated* • *Initial regimen* • *InitiationBaselineCD4Count*				
Clinical data • *Last encounter* • *Next Appointment* • *TB data* • *Cervical cancer data(women)* • *STI screening data*				

Refill data • *Last refill* • *Next refill* • *Regimen switches* • *Current regimen*				
Lab test data • *Viral load* • *CD4* • *Full blood count* • *LFT'S* • *Creatinine*				

Appendix 4: Facility DQI tool

#	Variable	File 1:		Consistency	File 2:		Consistency	File 3:		Consistency
		EMR	Patient file		EMR	Patient file		EMR	Patient file	
1	Date of Birth			TRUE			TRUE			TRUE
2	Date Confirmed HIV+			TRUE			TRUE			TRUE
3	Date enrolled on ART			TRUE			TRUE			TRUE
4	Date of ART Start			TRUE			TRUE			TRUE
5	Initial ART Regimen			TRUE			TRUE			TRUE
6	Current ART Regimen			TRUE			TRUE			TRUE
7	TB Screening at last visit (Done/Not done)			TRUE			TRUE			TRUE
8	TB Screening outcomes [e.g. No Signs, Pres. TB, on ART]			TRUE			TRUE			TRUE
9	IPT Start date			TRUE			TRUE			TRUE
10	IPT Status/outcome [e.g. completed, discontinued]			TRUE			TRUE			TRUE
11	IPT Outcome date			TRUE			TRUE			TRUE
12	2nd last VL Result [C/ml]			TRUE			TRUE			TRUE
13	2nd last VL date			TRUE			TRUE			TRUE
14	Most recent VL Result [C/ml]			TRUE			TRUE			TRUE
15	Most recent VL date			TRUE			TRUE			TRUE
16	Last Clinical Appointment date			TRUE			TRUE			TRUE
17	Next Appointment date			TRUE			TRUE			TRUE
EMR vs Patient file Consistency (%)				100%			100%			100%

Appendix 5: Indicator – DQI tool

Assessment Area		Indicator	EMR	Primary Source	CBS data set	DHIS	DATIM
HTS		Total Number Tested HIV					
		Total Receiving HIV + Results					
PMTCT	ANC	ANC Testing for HIV - Initial					
		ANC Positive (New) to HIV Test					
		Known positive (KP) status (at entry into ANC)					
		Number of pregnant women receiving HAART at ANC					
		Issued in ANC (Infant ARV prophylaxis)					
	L&D	Number of women Testing for HIV in Maternity- Initial					
		Testing HIV Positive (New) at Maternity					
		Number of women receiving HAART in maternity					
		Infant issued with ARV prophylaxis in L & D					
	PNC	Postnatal (<=6wks) Testing for HIV					
		Postnatal (<=6wks) Positive to HIV Test					
		Number of women receiving prophylaxis at post-natal care					
		Number of Infant receiving prophylaxis at post-natal care					
	EID	Initial PCR done (April-Aug. 2019)					
		Total PCR Confirmed Positive (April-Aug. 2019)					
		Infants Started on ART (April-Aug. 2019)					
ART		Number of adults and children starting ART					
		Number of adults and children Currently on ART					
PrEP		Total clients initiated on Prep					
		Total clients Currently on Prep					
IPT		Number started on IPT					