Standard Operating Procedures

Data Extraction 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 Extraction SOP for the Botswana HIV Case Based Surveillance

2.0 PURPOSE

2.1 To define procedures for the extraction of data collected during the implementation of the HIV Case-Based Surveillance (CBS) in Botswana.

3.0 SCOPE

3.1 This standard operating procedure applies to all who have direct involvement in data extraction during the implementation of CBS activities. This will include National Data Warehouse (NDW) officers, Monitoring & Evaluation (M&E) officers and CBS TWG members.

4.0 INTENDED USERS

National Level Staff

Ministry of Health and Wellness (MOHW) program officers and technical partners involved in the implementation of the HIV CBS activities, including data extraction, will have the following roles to play:

4.1.1 *National Datawarehouse (NDW) Administrators and Developers*

This team consists of database administrators and software developers, who are responsible for ensuring the extraction of required data elements from NDW into the CBS database. Create a CBS dataset for analysis by the data analysis team. Ensure that data is being pushed into the CBS database monthly, or per agreed timeline. Collaborate with the analysts and M&E team to implement and develop dashboards for data reviews at site and national levels.

4.1.2 *National Datawarehouse M&E Officers*

Consists of M&E officers from both the MOHW and Implementing Partners (IP) who are responsible for reviewing CBS data on a monthly basis for: (1) completeness, (2)timeliness, and

(3) other relevant data quality issues. Use CBS dataset to produce frequency tables report for data completeness and data validity, and provide feedback to sites on data quality gaps. The M&E officers will also ensure the timeliness of data submissions both from the HIV healthcare facilities and the generation of the CBS dataset. Representativeness or availability of case data (i.e., are all health facilities submitting data?) and prioritize follow-up activities for health facilities. These activities will not only be limited to CBS requirements, but will be used for other programmes as needed.

4.1.3 *Data Analysis Team*

This team consists of the M&E officers, MOHW surveillance officers and IPs data analysis support in the form of data analysts, biostatisticians and epidemiologists. They will be responsible for guiding the development of the CBS data analysis plan, design of analysis dataset, design and provide guidance to NDW administrators on dashboards. This team will prepare and submit quarterly and annual CBS reports to the technical working group and to identified stakeholders.

4.1.4 *CBS 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.

5.0 PROCEDURES

5.1 Backup Generation and Transportation

Backup generation is required for all sites that use the PIMS EMR system.

- For the PIMS EMR facilities that **are connected to the Government Data Network (GDN)**, data collected from these facilities is automatically backed up through the Backup Master to the ownCloud server. Data are transmitted on a monthly basis through the ownCloud client software installed on PIMS servers to the NDW.
- Facilities that **are not connected to the GDN**, the backup from the Backup Master will be physically collected monthly by the MOHW IT officers and transported to facilities that are connected to the GDN for upload to the ownCloud server.
- Facilities that do not have an existing EMR system will use a case report form (CRF). The CRF, once completed, will be entered to the CBS DHIS2 instance. (See Appendix A and B for the CRF)

5.2 Backup Restoration and Data Pulling to NDW

Facilities that use the Integrated Patient Management System (IPMS), a centralized system that links these facilities, will have their data pulled into the NDW daily. For facilities that submit backups to the ownCloud server, the NDW administrators pull the data from the backups as they are uploaded to the ownCloud server. This process should be completed by the third week of the month. CRF data from the CBS DHIS2 instance is to be pulled to the NDW monthly.

5.3 Data Extraction from NDW into the CBS Dataset

The NDW administrators and software developers are responsible for extraction of data from the NDW into the CBS dataset, and for producing reports that will contain details of the counts, including disaggregation by facilities, districts, sex age etc., for the entire dataset and merged dataset. Generating a report on the deduplication methodology applied, showing the number of records before deduplication and the result after deduplication including disaggregation; this will be completed with guidance from the analysis team. The following steps will be conducted monthly for backups that were received in the previous month, to produce a CBS dataset for the data managers, M&E officers, and data analysts.

- The data from the data sources are merged at the NDW. This merging process involves appending the records from PIMS, IPMS, and the CRF data from the CBS DHIS2 instance, into a single dataset.
- Merged dataset will then be deduplicated using the omang number (national ID) for records that have an omang and a pseudo generated ID number using name, sex, and date of birth for the records that do not have an omang number. This method is further explained in the record linkage section.

Deduplicated datasets will be de-identified (remove personal identifying information) at least quarterly and data will be used for visualization.

Deduplication Steps:

Steps:

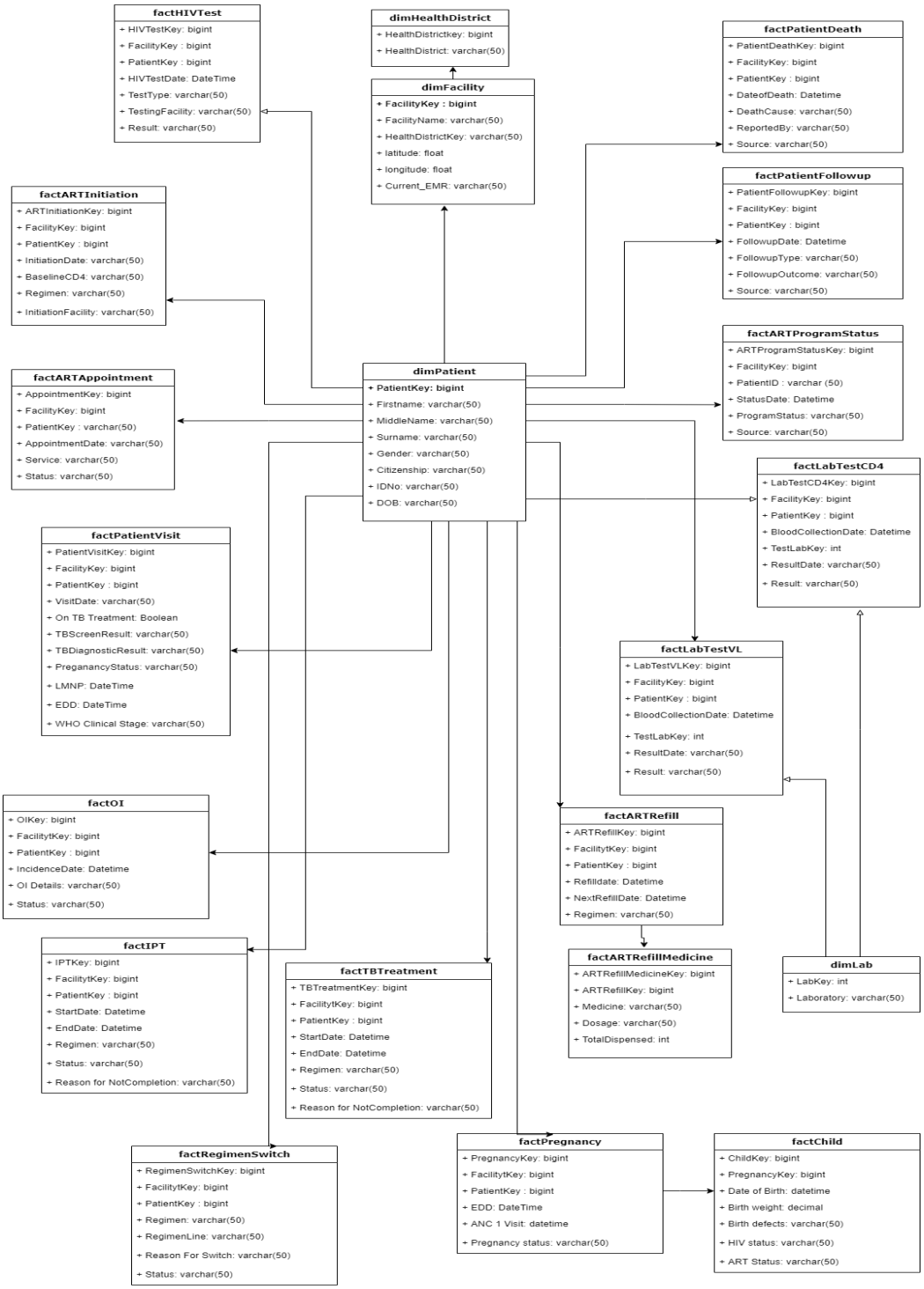
1. Generate a system unique identity (SUID) for all records by combining the firstname, surname, date of birth (DoB), and gender for all records.
2. Get valid Omang details where it is available. The Omang number is a unique national ID issued to citizens by the government of Botswana. It is a nine digit number with the fifth number as either 1 or 2, for males and females, respectively.
3. Remove any special character (#*\%/><) that may appear as part of an ID variable. Trim any white space.
4. Get passport numbers for noncitizens, check if the patient is a citizen and if the value under the passport variable is more than seven characters.
5. If more than seven characters, then step 4 applies, if not then proceed to step 6.
6. To get a unique identification for records without a valid ID: check the subset data produced from steps 1 to 4 that do not have either a valid ID or passport. Look through the dataset if there is any matching SUID to any other SUID that has a valid ID or passport. The ID or passport from the matching record is used as a valid unique id for the record.
7. Assign SUID as computed ID number: The remaining dataset without an ID is assigned the generated SUID number.

6.0 The CBS Dataset

The CBS analytical data set will be generated from the NDW and will cover all sentinel events that are needed for the CBS. These data elements are also provided in the Data Analysis SOPs. Data comes into the NDW from HIV healthcare facilities in a variety of ways already described in the Protocols for Establishing Case Based Surveillance in Botswana. Data elements required for CBS are then pushed into a staging CBS database within the NDW. This final clean, deduplicated, and deidentified dataset will then be extracted into analysis and visualization applications such as DHIS2 or PowerBI with clinical events for all patients represented longitudinally. The process for generating this data is as follows:

1. Unique patient records are pulled from the NDW into a CBS staging database within the NDW.
2. For each of the unique records, clinical events of interest (HIV CBS sentinel events, which includes testing records, appointment records, laboratory records etc.) will be pulled for a record to a separate database for each clinical event and also placed in the CBS staging database.
3. A Surveillance ID will be created for all records, to link them together across all event databases in the staging database.
4. The records are then combined to form a single and final CBS dataset that is ready for analysis.
5. The final CBS dataset will then be extracted from the staging database to the different analysis and visualization applications quarterly.

7.0 CBS Database Structure



Appendix A: Case Report Form for Newly Diagnosed Persons Living with HIV Infection.

Botswana Ministry of Health and Wellness

Case Report Form for Newly Diagnosed Persons Living with HIV infection

This form may be completed by a health care provider on the day of HIV diagnosis or within 7 days.

Please record all date as dd/mm/yyyy

Section A: Client Unique Identifier

(Personal identifying information from this section will not be included in the surveillance database)

1. **Name:** First_____ Surname _____ Middle _____
2. **Alternative name:** _____ Not available ☐
3. **Place of birth:** City/town/village_____ province/state/district _____ Country/code _____
4. **DOB**: / /
5. **National identification** (check all that apply):
   National ID (OMANG) : _____ Birth certificate number: _____
   Driver licence : _____ Passport number : _____ Not available ☐
6. **HIV care/ART ID number:** _____ Not available ☐
7. **Index case number**: _____ Not available ☐

Section B: Client Demographic Information

1. **Marital status:** Never married/single ☐ Married-monogamous ☐ Co-habiting ☐ Married-polygamous ☐ Divorced/separated ☐ Widow/Widower ☐
2. **Location and type of residence:**
   District name: _____ Village/ward/kgotla: _____
   Residence (check one only): House/apartment/flat ☐ Prison ☐ Temporary house ☐ Homeless ☐ Shelter ☐
3. **Age**: _____ year(s) _____ month(s)
4. **Sex at birth:** Male ☐ Female ☐

Section C: Facility information

1. **District:** _____ 2. **Testing Site Name/Code:** _____
5. **Reporting Site/Code:** _____
6. **Point of HIV testing service where the case was diagnosed:**
   HTS ☐ ANC ☐ Maternity ☐ IDCC ☐ TB ☐ VMMC ☐
   A&E ☐ Inpatient Ward ☐ STI ☐ OPD ☐ Child Welfare ☐
   Other ☐, specify:_____

Section D: Report reception/Data management information

1. Date Form Completed: / /
2. Date Report Received: / /
3. Date Report Entered: / /

Section E: Index Testing

1. **Contact of index case?** Yes ☐ Index case ID number _____ Not available ☐
   Index case ART ID number _____ Not available ☐
   No ☐ Refuse ☐

(If client aged < 5 years and identified through PMTCT: HTS and ART ID number of the biological mother should be used)

Section F: HIV Testing

1. **Date of first HIV positive test (dd/mm/yyyy):** / / **Type of HIV test:** ☐ Rapid Test ☐ PCR (EID testing)
2. **Date of HIV positive verification test (dd/mm/yyyy):** / /

Section G: Client Clinical History Information (client aged ≤ 5 years)

1. **Birth weight:** ____kg **Gestation at birth:** ____ weeks
2. **Maternal ART:** Yes ☐ No ☐ Don't know ☐
   If yes, ART initiation: Before pregnancy ☐ During pregnancy ☐ During birth ☐
   After giving birth ☐ Don't know ☐
   ART regimens taken before or during pregnancy or during or after giving birth (list all): _____
3. **Infant ARV prophylaxis:** Yes ☐ No ☐ Don't know ☐
   If yes, NVP ☐ NVP & AZT ☐ Other ☐ specify:_____ Duration:____weeks
4. **Birth defects (ICD-10):** Yes ☐ No ☐ Don't know ☐
   If yes, specify_____

Section H: Client Clinical History Information (client aged > 5 years)

1. **Date of most recent HIV-negative test:** / / or ____ months ago Never been tested ☐
2. **Ever been on PREP** Yes ☐ No ☐ Refuse ☐ Unknown ☐
3. **Ever been on ART** Yes ☐ No ☐ Refuse ☐ Unknown ☐
4. **Ever received ARV/ART prophylaxis to prevent mother to child HIV transmission?**
   Yes ☐ No ☐ Refuse ☐ Unknown ☐
5. **If the client is a girl/woman ≥12 years of age**: a. Pregnant ☐, *Gestation (weeks)*: ____
   b. Breastfeeding ☐, Post-delivery (months): ____ (Up to 24 months)
   c. Not pregnant or breastfeeding ☐

Section I: Client Clinical Information at the time of HIV Diagnosis

1. **Was the WHO clinical stage assessed?**
   No ☐ Yes ☐ → date: / /
   Result: Stage I ☐
   Stage II ☐
   Stage III ☐
   Stage IV ☐

2. **1st CD4:** Not done: ☐
   Sample collection date: / /
   Sample test date: / /
   Result count: ____
   Result percent: ____

3. **ART initiation**
   No ☐ → Referral ☐ Refused ☐
   Yes ☐ Initiation date: / /
   Regimen: _____

4. **Was cryptococcal infection diagnosed?**
   No ☐ Not done ☐
   Yes ☐ → date / /
   Results: CrAg positive ☐
   CM/disseminated ☐

5. **Was Tuberculosis diagnosed?**
   No: ☐ → Was TB preventive therapy (TPT) given?
   Yes ☐ → date: / /
   Regimen: _____
   (e.g., INH, 3HP)
   No ☐ → Reason: ☐ Client refused
   ☐ Contraindication
   ☐ Already completed TPT
   ☐ No drug supply
   Yes: ☐ → date / /
   Symptom screening positive ☐
   Sputum positive ☐ Xray positive ☐
   Other test positive ☐, specify _____
   Was the patient on TB treatment?
   Yes ☐ → Start date: / /
   Regimen: _____
   No ☐ → Why? _____
   Not done: ☐

6. **Were any other opportunistic infections diagnosed?**
   No ☐
   Yes ☐ → date / /
   If yes, specify _____

Appendix B: Case Report Form for Sentinel Events for a Previously Reported Case

Botswana Ministry of Health and Wellness

Case Report Form for Sentinel Events for a Previously Reported Case

This form may be completed by a health care provider every 3 months (or 6 months) from the date of HIV diagnosis per country guidelines. Please record all date as dd/mm/yyyy

Section A: Client Unique Identifier/Client Profile

(Personal identifying information from this section will not be included in the surveillance data repository)

1. **Name:** First_____ Surname _____ Middle _____
2. **Alternative name:** _____ Not available ☐
3. **Place of birth:** City/town/village_____ province/state/district _____ Country/code _____
4. **DOB:** / / **Age:** year(s) month (s)
5. **Sex at birth:** Male ☐ Female ☐
6. **National identification** (check all that apply):
   National ID (Omang): _____ Birth certificate number: _____
   Driver's license: _____ Passport number: _____ Not available ☐
7. **HIV care/ART ID number:** _____ Not available ☐

SECTION B: Care and Treatment Facility Information

1. **District:** _____
2. **Village/city:** _____
3. **Care and Treatment Facility Name/Code:** _____

SECTION C: Report Reception/Data Management Information

1. **District:** _____
2. **Reporting facility/Code:** _____
3. **Date Form Completed:** / /
4. **Date Report Received:** / /
5. **Date Report Entered:** / /

Section D: Client Latest Demographic Information

1. **Location and Type of Resident**
   a. District name: _____
   b. Village/ward/kgotla: _____

2. **Marital Status**
   Never married/single ☐ Married-monogamous ☐
   Co-habiting ☐ Married-polygamous ☐
   Divorced/separated ☐ Widow/Widower ☐

Section D: Client Testing and Treatment History

1. **Date of HIV diagnosis:** / / **Testing site name/code:** _____
2. **Date first enrolled in care/treatment:** / / **Care/treatment clinic name/code:** _____
3. **Transferred in:**
   No ☐, skip to Section E
   Unknown ☐, skip to Section E
   Yes ☐, →Question 4
4. **Transfer in date:** / /
5. **Previous facility name/code:** _____
6. **Patient was on ART:** No ☐ Unknown ☐
   Yes ☐, HIV Care/Treatment ID number _____
7. **Date of ART initiation in previous facility:** / / Unknown ☐
8. **Regimen** _____ Unknown ☐

Section E: Antiretroviral Treatment (ART) during the reporting period

Current **HIV Care/Treatment ID number:** _____ (Note: a patient could receive more than one regimen)

On 1st line: ☐ Date started on 1st line / / Regimen: _____
2nd line: ☐ Date started on 2nd line / / Regimen: _____
3rd line: ☐ Date started on 3rd line / / Regimen: _____

On a special (prescribed by a doctor) regimen (not 1st or 2nd or 3rd line):
No ☐ Yes ☐ → Date started on the special regimen: / / Regimen: _____

Reasons to switch a new ART regimen:
Treatment failure (Viral load not suppressed or drug resistance) ☐ Adverse drug reaction ☐ Gastrointestinal ☐
Skin ☐ CNS ☐ Haematological ☐ Hepatic dysfunction ☐ Metabolic ☐ Headache ☐
Kidney dysfunction ☐ Bone dysfunction ☐ Fatigue ☐ Treatment guideline change ☐
ARV shortage/stockout ☐ Drug-drug interactions ☐ Pregnancy/planning to become pregnant ☐
Other ☐, Specify: _____

Section F: Women and Child Health during the reporting period (for female patient only)

1. Was she pregnant?
   - Yes ☐ →
   - No ☐, skip to G
   - Don't know ☐, skip to G

a. Last day of menstrual period / / Don't know ☐
b. Due date / / Don't know ☐
c. Attend antenatal care? No ☐ Don't know ☐
Yes ☐ → Date of 1st ANC visit: / / Gestation (weeks): ___

2. Did she give birth?
   - Yes ☐, →
   - No:
     Miscarriage ☐, skip to G
     Stillbirth ☐, skip to G
     Abortion ☐, skip to G
     Don't know ☐, skip to G

The child's date of birth: / / Don't know ☐
Gestation at delivery: ___ weeks **Birth weight:** ___ kg
Birth defects (ICD-10): Yes ☐ No ☐ Don't know ☐
If yes, Specify _____
Was the baby diagnosed with HIV?
- No ☐ Don't know ☐
- Yes ☐ → Date of the diagnosis / / Not available ☐
  Was the child initiated on ART? No ☐ Yes ☐
  ART ID number? _____ _ Not available ☐

Section G: Laboratory Test Information

CD4 T cell count and percentage during the reporting period

Date of sample collection: / / CD4 count _____ cells/µL CD4 percentage _____%
Date of sample collection: / / CD4 count _____ cells/µL CD4 percentage _____%

HIV viral load RNA test during the reporting period

Date of sample collection: / / Detectable copies _____ copies/mL
Detectable copies _____ log Undetectable ☐

Date of sample collection: / / Detectable copies _____ copies/mL
Detectable copies _____ log Undetectable ☐

HIV Drug Resistance during the reporting period

Date of sample collection: / / Date of sample tested: / / Sample rejected ☐ Not done ☐
Major mutation results: NRTI ☐ NNRTI ☐ PI ☐ INI ☐ Other ARV class ☐, specify ART name_____

Section H: Clinical Information During the Reporting Period

Was Tuberculosis diagnosed?
No ☐ Not done ☐
Yes ☐ → date / /
Symptom screening positive ☐
Sputum positive ☐
Xray positive ☐
Other positive ☐,
specify _____
Was the patient on TB treatment?
Yes ☐ → Start date: / /
Regimen: _____
No ☐ → Why? _____

Was TB preventive treatment (TPT) completed?
No ☐ → Why? Non-adherence ☐
Developed active TB ☐
No drug supply ☐
Adverse drug reaction ☐,
specify _____
(e.g., rash, neuropathy, liver toxicity)
Yes ☐ → Start date: / /
End date: / /
Regimen: _____
(e.g., INH, 3HP)

Was the patient LTFU?
No ☐
Yes ☐ → date / /

Was the patient transferred out?
No ☐
Yes ☐ → date / /

Did the patient die?
No ☐
Yes ☐ → date / /
Cause of death: _____

Was cryptococcal infection diagnosed?
No ☐ Not done ☐ Yes ☐ → date / /
Results: CrAg positive ☐ CM/disseminated ☐
Was the patient on treatment? No ☐ → Why? _____
Yes ☐ → Start date: / / Regimen: _____

Was the WHO clinical stage assessed?
No ☐ Yes ☐ → date: / /
Result: Stage I ☐ Stage II ☐
Stage III ☐ Stage IV ☐

Were any other opportunistic infections diagnosed? No ☐ Yes ☐ → date / /
Specify: _____