Intervention Protocol Data Manual

Table of Contents

1.0 Background

The Botswana Combination Prevention Project (BCPP) was a research project conducted by the Botswana Ministry of Health (MOH), Harvard School of Public Health/Botswana Harvard AIDS Institute Partnership (BHP), and the U.S. Centers for Disease Control and Prevention (CDC). BCPP was a community randomized trial that examined the impact of prevention interventions on HIV incidence in 15 intervention and 15 control community. The interventions included extensive HIV testing, linkage to care, and universal treatment services. To reduce HIV incidence in the intervention communities, the UNAIDS 90-90-90 goals were used: 90% of HIV-positive persons know their status; 90% of persons who know status are to be on ART; 90% of persons on ART are to be virally suppressed.

2.0 Purpose

The purpose of this document is to provide overall guidance for the use of the BCPP public release datasets for the Intervention Protocol. Using the BCPP Research Database (RDB) that combines all study sources, analytical datasets were derived to support the analysis needs for the BCPP's study objectives and future use.

3.0 Definitions and Abbreviations

- BCPP: Botswana Combination Prevention Project
- BHP: Harvard School of Public Health/Botswana Harvard AIDS Institute Partnership
- CDC: U.S. Centers for Disease Control and Prevention
- CPC: Combination prevention communities, also referred to as the intervention communities, received the prevention interventions.
- EMR: Electronic medical records are digital forms that reflect patient care received at a clinic.
- HTC: HIV Testing and Counselling campaigns were done in the intervention communities. The acronym HTC will be used to represent this activity as a whole.

- IPMS: Integrated Patient Management System is one of two patient management systems used by the Botswana Ministry of Health to manage patient electronic medical records (EMR).
- MOH: Botswana Ministry of Health
- PIMS: Patient Information Management System II is one of two patient management systems used by the Botswana Ministry of Health to manage patient electronic medical records (EMR).
- RDB: The BCPP Research Database (RDB) was the central database created as part of the BCPP project to house all of the data collected for the research study. The RDB contains data from PIMS, IPMS, BHP and other sources. It was maintained daily by automated process to load and consolidate data.
- SAS: Analysis software used for statistical analysis.

4.0 Data Description

Personal Identifying Information (PII)

- The personal identifier, omang, is not available in the analytical data.
- The encrypted personal identifier, omang, is not available in the analytical data.
- The EMR patient registration numbers are not available in the analytical data.
- Event dates are not available in the public release data.
- A unique de-identified study identifier (i.e. bcppid) is available in the data for person-level analysis.

Data Format and Structure

- A prefix has been added to each dataset to identify that the data is associated with the study (i.e. BCPP).
- Available file types: csv

Analytical Datasets

- The following table lists the public release analytical datasets with an overview description.

Table 1. BCPP Intervention Public Release Datasets

Dataset Name	File Name	Dataset Description
HTC Enumeration Plot	▪ bcppr1_cpc_plotlist_open.csv ▪ bcppr1_cpc_plotlist_open_dictionary.rtf	▪ Round 1 plots identified in the communities and assessed by HTC. Includes the Evaluation Protocol 20% sample and Intervention Protocol covering the remaining 80% in the community.
HTC Enumeration HH	▪ bcppr1_htc_household_open.csv ▪ bcppr1_htc_household_open_dictionary.rtf	▪ Round 1 households identified on residential and habitable plots, whether or not the household was enumerated. ▪ Refer to BCPP Intervention HTC HH Enumeration Questionnaire to reference the data collection form.
HTC Enumeration Member	▪ bcppr1_htc_member_open.csv ▪ bcppr1_htc_member_open_dictionary.rtf	▪ Round 1 members enumerated in the visited households. ▪ Refer to BCPP Intervention HTC HH Enumeration Questionnaire to reference the data collection form.
BCPP HIV Testing and Counselling (HTC) Intake Forms	▪ bcpp_htcintake_open.csv ▪ bcpp_htcintake_open_dictionary.rtf	▪ Mobile and home-based testing through HTC in the CPCs ▪ All HTC intake forms for those who tested or refused during home or mobile testing. ▪ Event-level dataset (i.e. multiple tests per person). ▪ Refer to BCPP Intervention HTC Individual Intake Questionnaire to reference the data collection form.
BCPP Participant Summary in Intervention Arm	▪ bcpp_summaryintervention_open.csv ▪ bcpp_summaryintervention_open_dictionary.rtf	▪ Participants in the intervention arm of BCPP ▪ Contains study outcomes relative to specific-study research questions. ▪ This dataset is derived from the business rule requirements document that outlines the algorithms that define each variable. ▪ Person-level dataset (i.e. 1 row per person). ▪ Only participants with a valid identifier (i.e. omang or passport) are included to allow for deduplication across all sources to represent the individual, not event.

5.0 Summary Data Algorithms: Using PIMS and IPMS

This section describes algorithms that were used in the "BCPP_SummaryIntervention" dataset. This is not comprehensive of all analytical variables defined using algorithms in the summary analysis tables and does not reflect all data sources used to derive each variable. The following algorithms described below are focused specifically on the 2 EMR systems, PIMS and IPMS, and serves to provide examples of how the study team approached multiple data sources to measure study outcomes.

The "BCPP Summary Intervention" is the primary dataset containing one record for each unique participant who was identified by the BCPP study team in the Intervention communities. This dataset summarizes participant-level data across multiple data sources over the period of the study time. The unique identifier was an omang or passport. If participants did not have a valid or known unique identifier, they were excluded from this dataset and subsequently from all analyses per protocol. The ability to uniquely identify an individual across systems is a strength of this data for analysis.

Event dates have been removed from the public release data. "Time to" day variables have replaced event dates.

Deduplication of participants across PIMS and IPMS

To uniquely identify a person across PIMS and IPMS, a deduplication process was used across systems.

- Join participants across systems via OMANG. Use data from the system containing the latest Regimen Refill Date or Latest Show-up Date with ART Service, unless the IPMS patient record is an incomplete lab only record, then always take the PIMS record. An incomplete IPMS record can be identified by the field dw.dimIPMSPatient.isLabOnly = true.
- An incomplete IPMS record will be a record that comes from the BCPP_30_Communities_Visits files and does not exist in the BCPPDemographics file.
- For CD4 and viral load lab values, data is merged between the systems based on the defined aggregate function. Example: for first viral load date, take the first viral load value across both PIMS and IPMS.
- For the combined HIV+ date, data is merged between the systems so that the minimum HIV+ date becomes the merged HIV+ date.

Positive HIV Status

The following table outlines, at a high-level, which data elements were looked across and used to define the positive HIV status of a person in a clinic in the BCPP communities using PIMS and IPMS.

Data Element Description
A participant in PIMS is classified as HIV+ if any of the following are true:
HIV+ Test Date in lab results
HIV+ Test in Pretest Counsel (SRH)
Viral Load Test with results
Program Status “On HAART”, “On TAP”
Ever on ART, using the derived ART Initiation status
Regimen Refill
A participant in IPMS is classified as HIV+ if any of the following are true:
HIV+ Test Date in the patient record
HIV+ Test in lab results
ART Status – On Therapy
Viral Load Test with results
Ever on ART, using the derived ART Initiation status
Regimen Refill
IPMS visit date with visit reason specifically related to ART activity (e.g. ARV Follow-up, ART initiation)

Relevant to the Positive HIV status variable: the HIV Positive Date is the first date a person was found to be HIV positive based on the rules for “HIV Positive Status”. The HIV Positive Clinic Name is the clinic associated with the “HIV Positive Status” rule that the HIV Positive Date is associated with.

ART Initiation

The following table outlines, at a high-level, which data elements were looked across and used to identify who had initiated ART in a clinic using PIMS and IPMS.

Data Element Description
A participant in PIMS is classified as initiating ART if any of the following are true:
Minimum ART Initiation Date in HAART Initiation Table
Minimum Program Status Date when status changes to "On HAART or On TAP"
Minimum Refill Date with ART Regimen in Pharmacy Table
First viral load results occurring on or after June 1, 2016
A participant in IPMS is classified as initiating ART if any of the following are true:
ART Initiation date from patient table (occurring after a specified time parameter)
Minimum program status date where artstatus = 'On Therapy'
Minimum Refill Date with ART Regimen in Prescription table
First viral load results occurring on or after June 1, 2016

Given that there are multiple activities occurring across different systems and/or clinics, a process was required for record selection using priority logic. The order of activity to define the ART initiation date is:

1. ART initiation date
2. Program status
3. Refill activity
4. Viral load result

By system, PIMS was ranked over IPMS, except for with viral loads where IPMS was ranked over PIMS.

Retained On ART

The following table outlines, at a high-level, which data elements were looked across and used to identify who has been retained on ART using PIMS and IPMS. Asofdate is the date of reference defined by analysis needs.

Data Element Description
A participant in PIMS is classified as retained on ART if:
Alive AND
Refill with ART regimen and refill date within 120 days of the [asofdate] OR
Last PIMS ART Initiation Date within 45 days of the [asofdate] OR
Last program status of "On HAART" or "On TAP" is within 45 days of the [asofdate] OR
Viral load with a result within 180 days of the [asofdate] OR
Clinic visit date is within 180 days of the [asofdate] AND PIMS program status of "On HAART" or "On TAP" exists before the [asofdate] and visit date is after this program status date OR IPMS program status of "On Therapy" exists before the [asofdate] and visit date is after this program status date OR Actual initiation date in PIMS/IPMS exists before the [asofdate]and the visit date is after this initiation date
A participant in IPMS is classified as retained on ART if:
Alive AND
Refill with ART regimen and refill date within 120 days of the [asofdate] OR
Last IPMS ART Initiation Date within 45 days of the [asofdate] OR
Last program status of "On therapy" is within 45 days of the [asofdate] OR
Viral load with a result within 180 days of the [asofdate] OR
Clinic visit date is within 180 days of the [asofdate] AND PIMS program status of "On HAART" or "On TAP" exists before the [asofdate] and visit date is after this program status date OR IPMS program status of "On Therapy" exists before the [asofdate] and visit date is after this program status date OR Actual initiation date in PIMS/IPMS exists before the [asofdate]and the visit date is after this initiation date

Last Viral Load

This algorithm selects the last viral load for a participant across PIMS and IPMS. There is overlap between viral load results in PIMS and IPMS. The overlap occurs because of the viral load testing occurring at the laboratories that use the IPMS system and after the results are sent back to the PIMS clinics, the clinics enter the results into PIMS for patient management. Because of this process, there can be discrepancies between results due to data entry error at the PIMS clinic. To address this issue, a lab priority order was established to break any 'ties' on the unique key (omang and lab result date):

1. IPMS with the order location name
2. PIMS

If the IPMS order location name is unknown, then the PIMS clinic name will be used. If there are concerns around the IPMS order location name, generally, the PIMS clinic name accurately reflects the participant's clinic location since there can be error associated with the IPMS order location name.

Data Element Description
In PIMS: ▪ LabProfile = Viral Load OR ▪ Lab Test = Viral Load result AND ▪ Maximum sample date (i.e. the last date)
In IPMS: ▪ TestName = "Viral Load Count" AND ▪ Maximum sample date (i.e. the last date)

Viral Suppression

This algorithm uses the "Last Viral Load" algorithmically defined. This algorithm addresses the question: Is the person virally suppressed at study end? Viral load results with the values of 'ND, NP, unk' are excluded from this analysis.

Data Element Description
Viral Suppression = Yes if ○ Last Viral load sample date is within 18 months of 01 January 2017 AND ○ Last viral load result <= 400
Viral Suppression = No if ○ Last Viral load sample date is within 18 months of 01 January 2017 AND ○ Last viral load result > 400
Viral Suppression = unk if ○ Last viral load as defined above does not exist

6.0 BCPP Study Details

The following table provide general knowledge around the study in regards to time to support the analytical datasets. Table 2 defines the study cut-off dates across each outcome of interest for the Intervention Protocol. The selected dates allowed for identified persons to have time to flow through the 90-90-90 cascade. These end dates are unique to CDC data and is not reflective of Evaluation Protocol.

Table 2. BCPP Intervention Protocol Study Outcome End Dates

Outcome	End Date	Description
Last HTC interview assessment	30-Sep-17	All participants from HTC with an interview date through September 30, 2017
Last BHS/AHS interview assessment	31-Dec-17	All participants from BHS/AHS with an interview date through December 31, 2017
Last linkage date	31-Mar-18	The last date that a participant could link is March 31, 2018.
Last ART initiation date	31-Mar-18	The last date that a participant could initiate ART is March 31, 2018.
Last viral load result	30-Jun-18	The last date that a participant could have a viral load result is June 30, 2018.
Last retained on ART	30-Jun-18	The last date that a participant is considered retained on ART is June 30, 2018.
Death Date	30-Jun-18	The last date of death considered for a participant is June 30, 2018.