MINISTRY OF HEALTH

HIV CASE-BASED SURVEILLANCE IN KENYA

Guidelines for National Implementation

June 2020

MINISTRY OF HEALTH

HIV CASE-BASED SURVEILLANCE IN KENYA

Guidelines for National Implementation

June 2020

Attribution of support: This publication was made possible by support from the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) through cooperative agreements U2GGH001520-01 and GH001953 from the U.S. Centers for Disease Control and Prevention (CDC), Division of Global HIV &TB (DGHT).

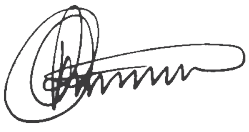
Recommended Citation: Ministry of Health, National HIV & STI Control Programme, HIV Case-Based Surveillance in Kenya. Guidelines for National Implementation. June 2020.

Foreword

The 2000 World Health Organisation (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS) guidelines for second-generation HIV surveillance recommended case-based surveillance (CBS) of HIV infection for settings such as Kenya; which have generalised HIV epidemics, widespread access to and uptake of HIV testing, access to effective care and treatment services, the capacity to collect, analyse and utilize case-based reporting data and the ability to ensure confidentiality of HIV information. The 2015 and 2017 WHO consolidated guidelines further recommended implementing CBS systems that collate data from different sources along the cascade of care and provided a common framework for reporting individual-level data to public health systems for surveillance and patient monitoring.

To align with these guidelines and the Kenya AIDS Strategic Framework (2014/2015-2018/2019), which calls for harmonized, timely and comprehensive routine HIV monitoring systems, it is time for Kenya to establish a HIV case-based surveillance system. Kenya already has an established Tuberculosis CBS system which the program relies on for programmatic monitoring, response planning and resource allocation. In the same vein, the Ministry of Health through the National AIDS and STI Control Programme (NASCOP) has put forward these guidelines to provide a pathway for the country's implementation of HIV CBS. Successful implementation of CBS will require the involvement, collaboration and support of national and county ministries of health, stakeholders and partners.

It is my hope that these guidelines will lead to a reliable long-term solution for HIV surveillance in the country. In particular, as Kenya works towards the achievement of UNAIDS 95-95-95 targets by 2030, we look forward to an efficient and flexible surveillance system that will provide critical information to guide HIV programme strategies, resource allocation and policy.



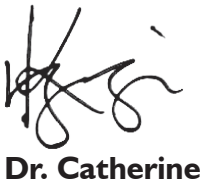
Dr. Pacifica Onyancha
Ag. Director
Directorate of Medical Services/ Preventive and Promotive Health
Ministry of Health, Kenya

Acknowledgements

The development and review of this document has been done as a collaborative effort of multiple institutions and stakeholders under the leadership of the Ministry of Health. We acknowledge with appreciation these institutions and organizations, both local and international, whose members spent many hours to ensure this document is developed and disseminated. We appreciate the efforts of the Ministry of Health Officials at the National AIDS and STI Control Programme (NASCOP), University of California, San Francisco (UCSF) and the US Centers for Disease Control and Prevention (CDC)-Kenya. Financial support for the development, review and publication of this document was provided by the US President's Emergency Plan for AIDS Relief (PEPFAR), through CDC-Kenya.

We wish to thank the individuals who contributed substantially to the development of this document: Joyce Wamicwe, Lilly Muthoni, Mary Mugambi, Dickens Onyango, Emily Zielinski-Gutierrez, Anthony Waruru, Peter W. Young, Eric-Jan Manders, Wanjiru Waruiru, Fitti Weissglas, Mwenda Gitonga, Diane Aluko, Mary Mwangome, Sandy Schwartz, Tim Kellogg, Joy Mirjahangir and Richelle Harklerode. In addition, we thank the DGHT CDC-Kenya Science office who reviewed this guidelines document.

Special appreciation to the team that reviewed and finalized this document: Mary Schmitz, Leonard Kingwara, Duncan Chege, Lydia Njuki, Samuel Kadivane, Violet Oramisi, Onesmus Muli, Stephen Mwangi, Margaret Wambere, Peter Mwangi, Lydia Mwangi, James Ransom, Benson Mwaniki, Ambrose Nyangeri, Vincent Ojwang, Maureen Akolo, Komu Mutunga, Charles Wachihi, Joseph Njunge, Margaret Ndisha, Joseph Barker, Davies Kimanga, Amee Schwitters, Rachael Joseph, Pascal Mwele, Jacob Odhiambo, Millicent Odongo and Olivia Ombuna. We also acknowledge colleagues from CDC-Atlanta Division of Global HIV and Tuberculosis for technical input and review of this document.



Dr. Catherine Ngugi
Head, National AIDS and STI Control Programme
Ministry of Health, Kenya

Table of Contents

Acronyms

AIS	AIDS Indicator Surveys
ANC	Antenatal Clinic
ART	Antiretroviral Therapy
CBS	Case-based Surveillance
CCC	Comprehensive Care Clinic
CDC	Centre for Disease Control and Prevention
COD	Cause of Death
CoE	Committee of Experts
DDSR	Division of Disease Surveillance and Response
KHIS	Kenya Health Information System
DHS	Demographic Health Surveys
DICE	Drop-in Centres
DSS	Disease Surveillance System
DQA	Data Quality Assessment
DWAPI	Data Warehouse Application Program Interface
DW	Data Warehouse
eCRF	Electronic Case Report Form
EID	Early Infant Diagnosis
EMR	Electronic Medical Record
FELTP	Field Epidemiology and Laboratory Training Program
FP	Family Planning
HEI	HIV-Exposed Infant
HIE	Health Information Exchange

HIS	Health Information Systems
HMIS	Health Management Information System
HTS	HIV Testing Services
IDSR	Integrated Disease Surveillance and Response
LTFU	Lost to Follow-up
MOH	Ministry of Health
MOU	Memorandum of Understanding
NASCOP	National AIDS and STI Control Program
NDW	National Data Warehouse
NHIF	National Hospital Insurance Fund
NUPI	National Unique Persons Identifier
PII	Personally Identifiable Information
PITC	Provider Initiated Testing and Counselling
PMTCT	Prevention of Mother-to-Child Transmission
SIRI	Strategic Information and Research Implementation
SOP	Standard Operating Procedures
STI	Sexually Transmitted Infection
TIBU	Treatment Information from Basic Unit
TB	Tuberculosis
TWG	Technical working group
UNAIDS	Joint United Nations Programme on HIV/AIDS
VCT	Voluntary Counselling and Testing
VL	Viral load
VMMC	Voluntary medical male circumcision
WHO	World Health Organization

Executive Summary

The recently released 2020 WHO Consolidated HIV Strategic Information Guidelines are an updated from previously released *2015 Consolidated Strategic Information Guidelines*, and an acknowledgment of the need for the world to continuously improve and build on methods and strategies for collection, analysis and use of HIV data at all levels to improve health sector response at all levels. Strategic information helps public health answer some critical questions about the epidemic: *Who is getting infected? Where are the infections? Who is on treatment? How are those on treatment doing? How is the programme performing in response to the epidemic?* One of the methods recommended by WHO for collection of routine strategic information is HIV case-based surveillance (CBS). HIV case-based surveillance is the individual-level reporting of diagnosis of HIV infection and defined sentinel events from every person diagnosed with HIV to a public health agency responsible for monitoring and controlling the epidemic.

Objectives: These guidelines present a framework for implementation of HIV CBS, and are adopted from WHO guidelines to the extent appropriate for the Kenyan context. They describe the setting up of an HIV CBS system in Kenya; including the components of the system, sources of the data to be reported to the system, the data repository, mechanisms of evaluating the system, how the system will be governed to ensure adherence to ethical principles, data analysis and dissemination of CBS data.

Audience: These guidelines are intended primarily to serve the needs of Ministry of Health's (MOH) HIV programme at the National and County levels. In addition, the guidelines serve the MOH partners engaged in the collection, analysis and use of HIV-related strategic information at all levels, including the at health facility.

Organization of the guidelines: These guidelines consist of the following 8 sections:

- Section 1-2: These sections provide the background of the HIV epidemic in Kenya and the definition, purpose and rationale for implementing HIV CBS in the Kenyan context. Finally the sections provide an outline of the coordination and implementation of HIV CBS in the country at all levels.
- Section 3-7: These sections provide an overview of the components of HIV CBS, sources of data and address issues of security and management of CBS data. Section 7 focuses on data analysis, interpretation and dissemination of CBS data.
- Section 8: This section gives a framework and includes measures for evaluating the CBS system.

1. Introduction to HIV Case-Based Surveillance

1.1 Background

In 2018, the prevalence of HIV in Kenya was estimated at 4.9% with an estimated 1.5 million adults and children infected with HIV nationwide (1). HIV prevalence varies with gender, with women experiencing higher prevalence than men (5.2% versus 4.5%) and geographic region (ranging from 21.0% and 20.7% in Siaya and Homabay counties of Nyanza region to approximately 0.1% in Wajir county in North Eastern region) (1). HIV disproportionately affects women more than men (prevalence estimates at 5.2%, and 4.9% respectively); and adolescent girls and young women between the ages 15-24 are particularly affected with their prevalence estimated at double that of their male counterparts (2.6% vs 1.3%) (1). In addition, key populations (KP) at risk for HIV contribute around 30% of new HIV infections(2). They have disproportionately higher HIV prevalence rates ranging from 29.3% among female sex workers (FSW), 18.2% among men who have sex with men (MSM) and 18.3% among people who inject drugs (PWID) (3).

There is a need to continually define and understand the demographic and behavioural characteristics of people living with HIV in Kenya; and monitor the dynamics and trends in the epidemic. To meet this need, Kenya needs to collect strategic information (SI) to inform policy and programme decisions which is crucial especially as the country seeks to implement tailored responses to an evolving epidemic. Surveillance systems are the cornerstone of the much needed SI for HIV programmes.

The HIV surveillance system in Kenya has been comprised of sentinel studies on HIV prevalence such as antenatal clinics (ANC) sentinel surveillance (1990-2011), periodic national population-based surveys such as the Kenya AIDS Indicator Surveys (KAIS) conducted in 2007 and 2012, the Kenya Population-Based HIV Impact Assessment (KENPHIA) in 2018, Demographic and Health Surveys (DHS) conducted in 2003, 2009 and 2014 and routine HIV programme data from health facilities. Although useful, these sources of surveillance data are infrequent, cross-sectional and, for programmatic data, are in aggregate form. It is therefore difficult to describe in a dynamic and timely manner the epidemic along the cascade of care from diagnosis, to entry into care, initiation on anti-retroviral therapy (ART), viral suppression and other outcomes. HIV case-based surveillance (CBS) can address these shortcomings. Furthermore, predictive modelling in 2014 determined that the HIV epidemic cannot be ended without providing treatment to the large majority of those who are infected, leading the UNAIDS to set the global target of 95-95-95 (4). These targets advocate that by 2030, 95% of all those infected with HIV should know their HIV status, 95% of those diagnosed with HIV are put on sustained antiretroviral treatment (ART) and 95% of those living with HIV and on ART will attain viral suppression (4). Monitoring of these targets can be accomplished through a CBS system since it is able to link individual records from diagnosis and the entire cascade hence providing accurate patient counts.

In recognition of the need for a CBS system, the National AIDS and STI Control Programme (NASCOP) piloted a CBS system in which individual-level, name-based, HIV data from adult and paediatric cases were retrospectively abstracted from patient files and reported to a CBS database. In the 2015 pilot, trained surveillance officers visited facilities, abstracted data and entered the cases into a pilot database. The pilot was conducted in a sample of 124 facilities in Kisumu and Siaya counties. The system utilized tablet devices and customized data collection software to capture and transmit data. The main findings from this 6-month pilot were:

- data to implement CBS in Kenya were available, but in varying formats and quality.
- systematically collecting, storing and analysing individual-level data was possible.
- the piloted surveillance officer-based approach was resource intensive
- a step-wise and multi-source approach method would need to be considered to roll out CBS at a national scale (5).

1.2 Purpose and Scope of Guidelines

To have a strong national HIV surveillance system, which is used to routinely describe the infected, their disease progression and where they are geographically. It is essential for an HIV programme to design programmes, allocate resources and address changes and needs as the epidemic evolves. To this end, an effective CBS system provides much richer information than the current system, beyond what is needed to track progress towards the 95-95-95 targets.

The purpose of this document is to provide guidance, options and examples for the design and implementation of an HIV CBS system, which can be designed to leverage existing systems that collect HIV-related data on a routine basis. The document is targeted to inform the design of a CBS system for public health use and focuses on HIV infected patients. These guidelines are written for use by all stakeholders in HIV surveillance, especially epidemiologists, programme managers, policy makers and public health informaticians.

In terms of scope, this document does not address program evaluation or patient care needs, hence it does not include tracking of individuals who are HIV-uninfected, even if they are at risk for acquiring HIV.

1.3 Case-Based Surveillance and Recency

In 2000, the World Health Organisation (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS) published guidelines for second-generation HIV surveillance to capture the distribution and trends of HIV within a population as accurately as was possible at the time. The guidelines recommended CBS or case-based reporting of HIV infection for settings such as Kenya, which are experiencing generalised HIV epidemics, have widespread access to and uptake of HIV testing, have access to effective care and treatment services, the capacity to collect, analyse and utilize case-based reporting data and have the ability to ensure confidentiality of HIV information (6).

HIV CBS involves the capture of individual-level information from persons diagnosed with HIV infection. Case-based surveillance can measure and characterise persons newly diagnosed with HIV, their immune status at diagnosis and provide information on the number and characteristics of persons living with HIV, the time from diagnosis to entry into care, retention in care, use of ART, time from ART initiation to viral suppression, the proportion of prevalent cases who are virally suppressed, and HIV mortality (6). Moreover, HIV CBS systems can monitor and detect shifts in the epidemic and enable characterisation of the population failing to be reached by programmatic efforts.

As Kenya approaches achievement of the UNAIDS 95-95-95 targets, the details of those who are most recently diagnosed and integrated into care become more critical to describe in order to inform the best strategies for the "end game" for the epidemic. Information on the newly diagnosed persons may include whether HIV infection is recent or not. Point of care tests (POC) for recent HIV infection may be used to diagnose recent HIV infection. These tests, similar to HIV rapid diagnostic tests, can be performed by trained health care providers within routine HIV testing services (HTS) and generate HIV recency results within minutes. Although results of recency testing do not alter HIV status designation or clinical care provided to newly HIV diagnosed individuals, they can be used to prioritise recent cases for index partner testing in order to increase HIV case detection rates and interrupt ongoing chains of HIV transmission networks. At the population level, routine epidemiological analysis of recency results can be conducted to monitor trends in recent infection and to identify hotspot locations and sub-populations with recent HIV infection to inform targeted interventions.

Case-based surveillance complements and validates data generated by other surveillance systems. It can be a source of information on infections among sub-populations missed in other surveillance systems including information on HIV among children. The WHO/UNAIDS second-generation guidelines likewise note the role that can be played by mortality reporting that includes HIV-related cause of death, which is integrated as a component of individual case-based reporting. The 2015 WHO consolidated SI guidelines recommend implementing CBS systems that collate data from different sources along the cascade of care (7). The collation of data from different sources can be done using unique and other matching processes using demographic data and available identifiers for HIV-infected patients to enable tracking of the epidemic and to show the impact of interventions over time (7). Further, the WHO 2017 consolidated guidelines for person-centred HIV monitoring and case-based reporting provide a common framework for reporting individual-level data to public health systems for surveillance and patient monitoring (8). These guidelines also include guidance on the information system components as well as components of a national CBS guidelines document such as this one. Some elements in this CBS guidelines document are adopted from the 2017 WHO guidelines to the extent appropriate for the Kenyan context.

1.4 Policy Support for CBS

The Kenya Ministry of Health (MOH) is currently implementing Integrated Disease Surveillance and Response (IDSR) that requires weekly reporting of notifiable diseases. New HIV cases are among the priority IDSR conditions reportable on a weekly basis (9). However, only aggregate cases are reported. Patient-level data that would have been useful for detailed analysis are excluded. The Division of Disease Surveillance and Response (DDSR) guidelines do not specify what additional information (beyond the reporting facility and the number of cases) is required to be included in the current IDSR report. Although CBS for specific epidemic-prone diseases like measles, polio, viral haemorrhagic fevers, is ongoing under IDSR, CBS for HIV has not been implemented. In 2012, CBS for tuberculosis (TB) was introduced based on the existing policy and legal framework. CBS for HIV will likewise be implemented within the 2019 Data Protection Act (10). However, as described below, the existing policy and legal framework will be institutionalized to ensure effective compliance with the reporting, confidentiality and data security requirements of the reporting system.

1.5 Current HIV Monitoring Tools

The MOH conducts routine programme monitoring that collects information on the number of persons testing for HIV, the number of new HIV diagnoses, the number of persons accessing HIV care and treatment services and other HIV-related indicators. In government health care facilities, patient data are collected at the individual level and recorded onto both individual patient records such as MOH 257 (the Clinical Encounter "Green Card") and MOH patient registers.

A description of the forms and the purpose for which the data are collected is provided in table 1. These will be the primary sources for CBS data.

Table 1: Ministry of Health Kenya's HIV related forms and registers for routine data collection*

MOH Form	Purpose	Where collected
MOH 257 Clinical Encounter Form	Patient encounter form for care and treatment services. Tracks HIV care and treatment services	Care and treatment service clinics at every encounter
MOH 366 HIV Daily Activity Register	Summarises daily care and treatment activities, i.e., number of people served for different services	Care and treatment service clinics
MOH 362 HTS Laboratory and Linkage Register	Registers people receiving HTS services. Monitors linkage by recording the comprehensive care clinic (CCC) number of those testing positive for HIV	At HIV testing sites
MOH 361 A Treatment Preparation Register	Records clients who are enrolled in care but not yet started on treatment. Assigns CCC number	Care and treatment clinics at enrolment
ART and Cohort Register MOH 361B	Records clients initiating ART. Each page lists those registered in that 1 month, i.e., a cohort	Care and treatment clinics at start of ART

MOH Form	Purpose	Where collected
MOH 405 Antenatal Register	Records all pregnant mothers and services they received per visit	At MCH
MOH 333 Maternity Register	Records all mothers coming into the facility in labour for delivery and services they received	Labour ward
MOH 406 Postnatal Register	Records all post-natal mothers who come to attend post-natal clinic and captures the services they received	Where post-natal services are offered, e.g., MCH or maternity ward
HIV Exposed Infant (HEI) Follow-up Card	Clinic-held card for every exposed infant captures services and is longitudinal. Filed together with mother's CCC file	MCH
HIV Exposed Infant Register	Records services provided to HEI per visit. Longitudinal for 24 months per HEI. Lists those born in that 1 month, i.e., a cohort	MCH
MOH 731 form Comprehensive Summary	Comprehensive summary for reporting HIV services. Completed monthly.	Facility based form

*Forms listed in table were most recently revised between May-July 2017

1.6 Surveillance Data Reporting and Repositories

Aggregate data from various registers are compiled and reported on HIV programme performance on a monthly basis onto the Comprehensive Summary form (MOH 731) at facility level, which is then abstracted by sub-county Health Records Officers (HRO) into the Kenya Health Information System (KHIS). Relevant to epidemic surveillance, the Comprehensive Summary Form (MOH 731) is used to monitor the number of people testing positive for HIV, the number started on ART and the number diagnosed with TB. Health facilities are also required to report aggregate numbers of new HIV cases on a weekly basis to DDSR as part of the IDSR strategy. Facilities that use electronic medical record (EMR) systems for managing HIV-infected patients also periodically submit individual-level clinical data for patients in their systems to the national data warehouse at NASCOP. A national individual-level viral load database is used to monitor and document all viral load results in the country (11). Kenya also has a paper-based Civil Registration and Vital Statistics (CRVS) system to document all deaths and their causes. In addition to routine reporting of surveillance data, Kenya has relied on antenatal clinic sentinel surveys, which were conducted bi-annually until 2011, and Demographic and Health Survey (DHS) and AIDS indicator surveys (AIS), currently referred to as population-level HIV impact assessment (PHIA), both conducted every five years. These data sources have been supplemented by bio-behavioural surveys among key populations, mortality surveillance and mathematical modelling.

1.7 Limitations of Current Surveillance Systems and Repositories

Currently, several different repositories exist for HIV data, and, while they offer various strengths, none meet the requirements of HIV CBS. While these repositories of data have been useful in decision making and strategic planning, they have several limitations:

- The DDSR database does not provide an adequate tool for CBS as it only captures aggregate counts of newly infected people living with HIV (PLHIV).

- Data in KHIS are aggregated, making it impossible to associate individual characteristics of persons and their outcomes. For example, it has a limited ability to describe those who test HIV positive. In addition, it is not possible to de-duplicate data that are aggregated; so over-estimation of the number of persons diagnosed and receiving care is likely to occur.

- Current data sources cannot be integrated for actual individual-level surveillance due to lack of a nationwide unique patient identification mechanism. While there exist some forms of uniquely identifying information such as National Identification (National ID) and National Health Insurance Fund (NHIF) numbers, these are not routinely documented at health facilities, and they have limitations of their own. Not all persons have a National ID, which is issued to Kenyans aged 18 years and older; NHIF cards are also limited to persons who have signed up for or pay for the insurance service. In 2019, Kenya introduced a unique identifier for all citizens (huduma number). It is envisioned that this number will be used in multiple settings to uniquely identify persons. However, the operationalization of this has not yet taken effect since policies to enable its use in a variety of settings including healthcare are not yet in place.

- Current data repositories that receive individual-level data such as the national data warehouse (NDW) and the DDSR database, do not include all HIV-diagnosed individuals.

- Data for services delivered in some private facilities are not captured in MOH reporting systems. As such, HIV positive individuals diagnosed and/or receiving services in private facilities will not always be reported.

- The current national civil registration and vital statistics (CRVS) system lacks standardization in cause of death coding, generally does not document HIV status at time of certification of cause of death, has aggregated data prior to reaching the national level and has a low estimated national coverage of death reporting (<50%) limiting the generalisability of statistics derived from the system (12), and has limited ability to link to HIV records.

- Periodic surveys, while useful for supplementing programmatic data, are limited in providing timely information on epidemic trends and are costly to undertake.

As with surveillance systems for other infectious diseases, the setting up of an HIV CBS system requires prior determination and description of the components of the system; sources of the data to be reported to the system, the data repository, mechanisms of evaluating the system, how the system will be governed to ensure adherence to ethical principles, and possible data analysis outputs.

This document describes how all these elements shall be conceptualized.

2. Coordination and Implementation Framework for HIV Case-Based Surveillance

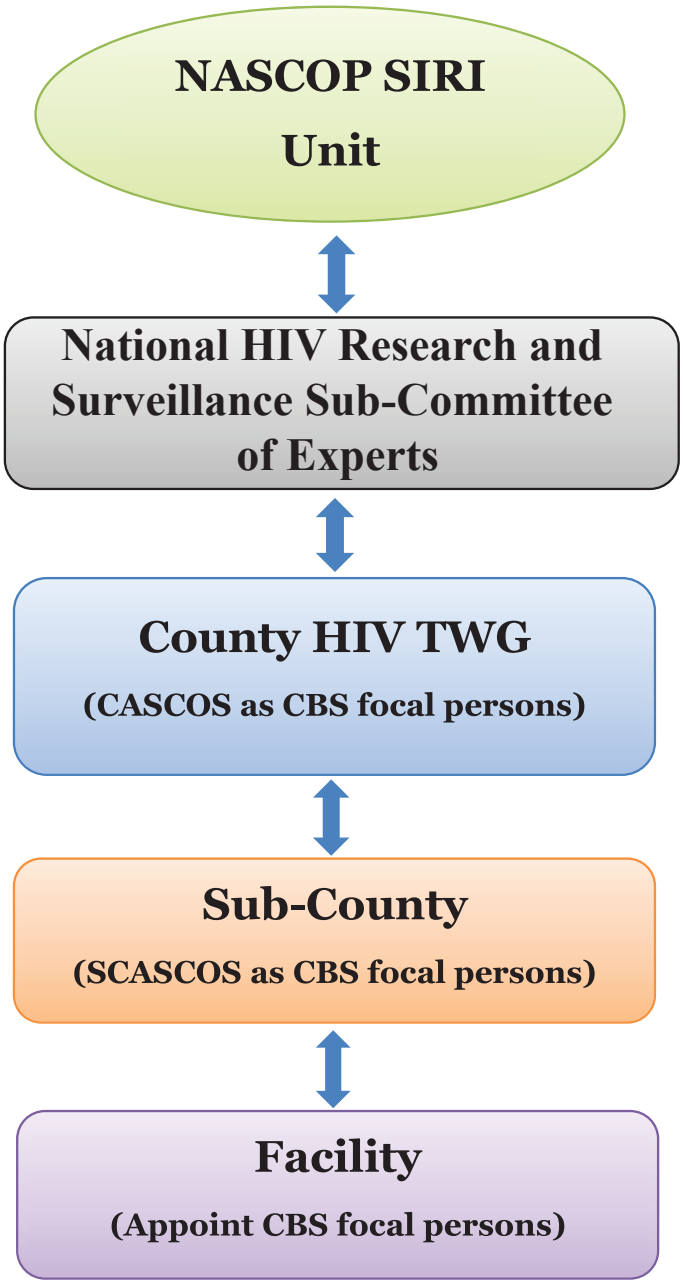
2.1 Coordination Mechanism

This section describes the mechanism for coordinating an HIV CBS system in Kenya. This coordination mechanism defines the roles and responsibilities of various stakeholders involved in the planning, implementation and utilization of CBS. The coordination mechanism will be crucial in engaging CBS stakeholders (funders, case reporting agencies, people living with HIV, civil organizations and users of the CBS data). The mechanism will also be important in securing buy-in and support of CBS system users.

There will be a two-tier coordination structure at National and County levels for the CBS system. This structure will coordinate the various components of the CBS system and support the use of the surveillance data for evidence-based decision making. At the national level, CBS coordination will be undertaken by the committee of experts (CoE) whose members are selected from diverse organizations including:

1. Surveillance lead in NASCOP Strategic Information and Research Implementation (SIRI) unit who is the chair of the CoE
2. Representative from DDSR
3. Representative from the Division of Health Management Information System (HMIS)
4. Representative from the Field Epidemiology and Laboratory Training (FELTP)
5. Representative of HIV service providers
6. National AIDS Control Council representative
7. Representatives of partners implementing HIV-related services
8. National Public Health Laboratories representative
9. Civil Registration and Vital Statistics representative
10. Representatives of PLHIV
11. Technical persons including epidemiologists, statisticians and health information system (HIS) specialists

Figure 1: Coordination Structure of HIV CBS system



At the county level, CBS coordination will be undertaken by county HIV technical working groups (TWGs). County HIV/AIDS coordinators (CASCOs) will serve as the county TWG focal person and work with the sub-county HIV/AIDS coordinators (SCASCOs) to implement CBS down to the facility levels (Figure 1). The roles and responsibilities of the different stakeholders involved in CBS implementation are outlined in Table 2.

Table 2: Roles and responsibilities for Case-Based Surveillance

Organization	Roles and Responsibilities
NACC	• Policy formulation and advocacy • Ensure data generated from CBS are disseminated and used to inform evidence-based decisions at national and community level • Mobilization of resources
NASCOP	• Develop and update guidelines on CBS • Provide technical oversight for the implementation of CBS in Kenya • Define reporting procedures and data sources • Obtain comprehensive look of the epidemic by compiling data from all data sources • Provide training and technical support to surveillance programs at sub-national levels • Provide material required for training and surveillance activities • Host and maintain a refined, complete, accurate and secure database • Provide the platform for data transmission to sub-national levels • Analyse, interpret, and disseminate routine reports on HIV infections in Kenya • Routinely evaluate the surveillance programme • Provide guidance on de-duplication of cases at sub-national levels • Coordinate national level data review and dissemination activities • Establish and coordinate the committee of experts to advise on developing, implementing, monitoring, evaluating and improving HIV surveillance systems in Kenya
County Health Departments	• Coordinate with NASCOP in the implementation of HIV CBS at county levels • Work with county-level collaborating partners to plan and implement case-based reporting and surveillance • Provide financial support for training of county staff • Provide county-level surveillance personnel to conduct routine data cleaning and data quality reviews • Provide support for field work (data collection, site visits, site supervision) of the county-level surveillance staff • In collaboration with MOH and partners, conduct county level data analysis and interpretation • Coordinate county level data review and dissemination activities on the HIV epidemic
HMIS in MOH	• Serve on the surveillance committee of experts • Facilitate HIV CBS is implemented at all facilities
HIV implementing partners in service provision	• Allocate resources to support implementation of CBS • Support capacity building for CBS • Participate in review and updates of guidelines for CBS

2.2. Capacity development for CBS

Capacity development of CBS stakeholders will be necessary for the successful implementation, operationalization and utilization of the CBS system. A package for sensitising and training stakeholders will be developed, reviewed and updated periodically by the surveillance CoE. Sensitising implementing partners and health leadership amongst other stakeholders will be coordinated by the NASCOP SIRI. Training and sensitising CBS stakeholders will be conducted both periodically and as need arises.

3. Components of HIV Case-Based Surveillance

The national CBS system aims to collect longitudinal data on children, adolescents and adults identified with HIV infection into a centralised national database. To do this, the system will utilise existing systems to routinely collect patient monitoring data at facilities and laboratories as well as vital statistics registration systems. The system is expected to be flexible enough to incorporate a variety of data collection systems and sources. The vision is to have a CBS system capable of characterising the HIV epidemic based on longitudinal collection of key clinical, virologic and immunological events in the natural history of the disease.

3.1 Case Definition

In the context of these guidelines, case definition is the process through which a HIV diagnosis is arrived at. Table 3 below highlights how a HIV case definition will be arrived at:

Table 3: Kenya case definition for HIV infection

Adults and children 18 months or older:

HIV infection is diagnosed based on:

Positive HIV antibody test (rapid or laboratory-based enzyme immunoassay). This is confirmed by a second HIV antibody test (rapid or laboratory-based enzyme immunoassay) relying on different antigens or of different operating characteristics;

and/or;

Positive virologic test for HIV or its components (HIV-RNA or HIV-DNA or ultrasensitive HIV p24 antigen) confirmed by a second virologic test obtained from a separate determination.

Children younger than 18 months:

HIV infection is diagnosed based on:

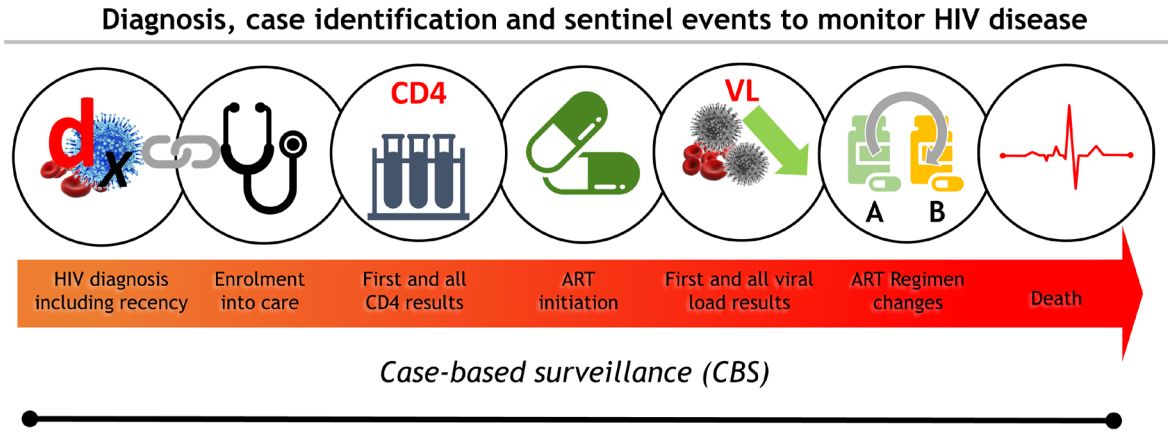
Positive virologic test for HIV or its components (HIV-RNA or HIV-DNA or ultrasensitive HIV p24 antigen) confirmed by a second virologic test obtained from a separate determination taken more than four weeks after birth.

**Positive HIV antibody testing is not recommended for definitive or confirmatory diagnosis of HIV infection in children until 18 months of age.*

3.2 Definition of Sentinel Events

Sentinel events, for the purposes of these guidelines, refer to HIV diagnosis, enrolment into care and initiation on and switch/change in ART regimen, CD4 count and viral load (VL) measurements, and death in HIV-infected individuals (Figure 2). Since many of the reportable sentinel events proposed for the CBS system are already collected by routine patient monitoring, CBS will use standard MOH monitoring tools to capture patient information on key sentinel events for each PLHIV (Figure 2).

Figure 2: Sentinel events in monitoring the spectrum of HIV disease



The data for these key events may be obtained from clinical, laboratory, and any other records deemed necessary. At a minimum, case reports should be reported or updated for HIV diagnosis, entry into care, ART initiation and death, if it occurs. In addition, all immunological and virologic data, results from initial CD4 and all VL tests results should be reported. These sentinel events are advised by the 2017 WHO case surveillance guidelines. The system is designed to receive and maintain data from a single individual obtained from multiple sources at different points in time. Using methods to uniquely identify the individual, new or additional sentinel events will be linked to an existing record and not entered as new case.

In monitoring of the 95-95-95 targets, the accurate and timely reporting of these sentinel events is important (4). To measure the first 95, the reported HIV diagnoses as a proportion of estimated PLHIV is computed. For the second 95, the proportion of those diagnosed with HIV that are initiated on ART is calculated. To compute the last 95, the number of cases reported to be virally suppressed as defined per the latest treatment guidelines out of those initiated on ART is calculated. All the computations should be done for a defined time period. Definitions of some events that may be captured in HIV case-based surveillance are shown in Table 4.

Table 4: Definitions of sentinel events for a minimum CBS data set

Sentinel event	Definition
First positive test indicative of HIV diagnosis	• Earliest date of HIV diagnosis determined according to the current national HIV testing algorithm.
Enrolment into care	• Date that any case of HIV is registered in clinical care; could be inferred by record of a CD4 test, VL test or ART initiation.
First CD4 test and all CD4 results	• The first CD4 test is the earliest CD4 percentage or count available and subsequent CD4 tests and their results. • The date that the specimen is taken is documented along with the result of the test.
Initiation of ART	• Date on which ART is first prescribed.
First and all VL results	• All VL results either baseline, routine or targeted follow up. • The date that the specimen is taken is documented along with the result of the test. • Results should include the exact copies/μL.
Change in ART regimen	• Date on which first-line ART regimen is changed to second-line. • Date when second-line regimen is changed to third-line.
Death	• Date of death reported in any case of HIV, regardless of the cause of death. • Underlying cause of death

In reporting a case, a **minimal set of variables** (beyond the sentinel event that triggered a report) will be required (Table 5). The purpose of ensuring these variables are included in each case report submission or update is to support case identification and matching of records. The minimum set of variables list can be updated to accommodate changes to data sources, policies and/or guidelines.

Table 5: Minimum variables required for reporting each sentinel event

Minimum variables to report with each case report for a sentinel event
• Reporting facility name and Master Facility List (MFL) code
• Date of report on case report form
• First, middle and surname of patient
• Patient identifier(s)
• Date of birth or age at diagnosis if birth date is unknown
• Sex

3.3 Case Report Form

A case report, according to the WHO 2017 consolidated guidelines for person-centred HIV monitoring and case reporting refers to both reports of new HIV cases and subsequent sentinel events related to existing cases (7). The HIV case-based surveillance form will collect demographic data, HIV diagnosis, clinical events, laboratory investigation results and patient status including but not limited to lost to follow up, death and transfer out. The case-based surveillance form in Annex I provides the current exhaustive list of variables that will be collected by the HIV CBS system. This form may be updated through the surveillance CoE to meet changing data requirements and to improve the surveillance system.

3.4 Unique Identification, Case Record Matching and Linkage

The accuracy of patient identification can be improved over time using several approaches until a robust unique identifier has been established nationwide. When utilising national identifiers that are not health-specific, several considerations should be made including: the ability to re-identify a citizen from the national identification number, thereby risking the loss of patient privacy; the real and perceived risks of identity fraud and its impact on the health sector; the acceptability of a national identifier for health by certain vulnerable and high risk groups of the population and the cost considerations of a parallel identification system for health (13).

In Kenya, there is currently no national unique personal identifier (NUPI) that is implemented in all sectors for all ages; neither is there a unique healthcare ID[1]. The lack of either makes it difficult to link individual-level data from different sources. However, efforts to implement existing policies around unique identification are ongoing. In the absence of such identifying documentation, Table 6 shows options of approaches in Kenya that could be used for unique identification and the matching and linkage of cases. Based on findings from the CBS pilot, probabilistic or score-based matching algorithms for case identification, matching and linkage performed better than deterministic approaches (14). Kenya should use probabilistic algorithms until such a time that better alternatives are available. This algorithm will require specific identifiers in order to facilitate linking the same person using the same highly specific pieces of information coming from different sources. Having functioning, complete and timely electronic management systems for patient data is critical to such linkage.

For facilities utilizing paper-based data systems, linkage of case reports presents unique challenges. For this reason, all records entered from paper-based sites should go through a staged de-duplication process before being received into the surveillance database. This means that these sites need basic algorithms for identifying cases that are already reported from that site and support de-duplication of cases locally at the site.

Standard operating procedures (SOPs) for routine identification and merging of duplicate case reports will describe the variables that must be collected for patient identification. In general,

1 *In February 2019, the Kenya government begun the roll out of Huduma Number. This is a number generated by the National Integrated Identity Management System (NIIMS) as the 'source of truth' of identity of all persons residing in Kenya. Unlike a national ID or birth certificate that is issued to only Kenyan Citizens, Huduma Number will be issued to all persons residing in Kenya.(hudumanamba.go.ke Accessed 04 03 2020) Plans will be made to operationalize the use of the Haduma Number in the Health Sector as CBS progresses.*

however, while the country should mainly use probabilistic approaches, it's the sample size of duplicates that warrants cross-validation and whether one or both deterministic and probabilistic approaches are to be used.

Table 6: Features for case identification, matching and linkage

Option	Key features
National identifier	A fixed ID that is issued by the government • Issued to all cases • Unique to one case • Works in online and offline scenarios • Can be used from HIV testing services (HTS) to care and treatment sites
Algorithm for matching	Attempts to match patients by comparing certain identifiers Could be probabilistic or deterministic **Probabilistic/score-based matching:** • Requires that certain identifiers are sent to a central data store • Identifiable variables must be kept in a central database • It is subject to not being able to resolve identity when demographic data are incomplete **Deterministic and fuzzy matching:** • Manageable with lower case loads • Lower match rates of cases, compared to probabilistic • Human resource intensive since human adjudication is an integral part

4. Data Sources

The national CBS system will utilize various data sources that capture information on HIV infected persons. These data sources include HIV testing sites, prevention of mother-to-child transmission (PMTCT) clinics, care and treatment clinics, laboratories and vital statistics registries (Table 7). Data collected on sentinel events will include HIV diagnosis, enrolment into care, initiation on ART, change of ART regimen, CD4 and VL measurements and death.

4.1 HIV Testing Services Sites

HTS sites vary in how they document HIV testing data. The four most common scenarios at facilities are as follows:

- **eMobile HTS (eHTS) at facilities with EMR.** In such facilities, HTS points of service utilize the eHTS application, which mirrors the MOH 362 paper register to enter testing services delivered. This application serves as an electronic MOH 362 and uploads all of its data into a designated EMR server. The HTS data are used in the EMR to support the continuation of care for HIV positive individuals.

- **Paper HTS at facilities with EMR.** In these facilities HTS service delivery is documented on paper registers (MOH 362). There is no client file/document linkage between the paper and the EMR.

- **eMobile HTS only.** Such facilities have implemented the eMobile HTS mobile application but continue to have paper-based clinical care documentation (green card/MOH 257).

- **All paper.** In these facilities, all service delivery is documented in paper-based MOH362 registers or files.

4.2 HIV Care and Treatment Facilities

The MOH Division of HIS has recognized the need to improve the use of information and communication technology in the health sector. One important objective is to have standardized and interoperable applications, including EMR systems.

EMRs are a key category of health information systems which are used for collection and hosting of HIV clinical information. In Kenya, most public facilities using EMR system use one of the three nationally accepted EMRs which adhere to the Kenya EMR guidelines: IQ-Care, Kenya EMR and Open MRS. In total, approximately 61% of all HIV patients in Kenya receive care services at a facility that uses a standard EMR, making them a rich source of case data.

The majority of health facilities however, do not use EMRs for general clinical care and management of patient records. An electronic data capture/entry solution/system will be made available depending on the facility needs. The system will be used for retrospectively transcribing case reports and uploading them to the surveillance database. Health facilities with paper records will

not be required to capture sentinel events beyond HIV diagnosis; however, should resources and technology allow, secure longitudinal data capture for facilities with paper-based reporting systems may also be implemented at selected sites in the future.

The system to be used for surveillance data capture and reporting at non-EMR sites will meet the following criteria:

1. Allow retrospective data capture of case reports and potentially longitudinal update of sentinel events
2. Allow off-line data entry
3. Have basic algorithms for identifying cases that are already reported from a non-EMR site and support de-duplication of cases locally at the site
4. Allow secure exchange of minimal set of demographic information through connectivity at intervals with the surveillance database
5. Meet the minimum requirements for data security as defined in chapter 7.

While eventually equipping all facilities with standard EMR systems is preferable in terms of overall benefit to the facility, it is also recognized that availability of electricity and network, and the remoteness of some facilities makes the use of EMR challenging.

4.3 Laboratories

Laboratory information systems are another source of CBS data. Laboratory-based surveillance systems are critical and complementary to notifiable diseases systems in the surveillance of infectious diseases and often are the primary source of surveillance data for many countries. Linked to the CBS database, laboratory information systems would add a critical layer of information for both validation and coverage of lab-related information in the CBS database. Furthermore, laboratory systems present a unique opportunity to capture HIV-infected individuals who receive services at paper-based sites. Where available, CBS will utilize electronic laboratory data from networks of regional laboratories that test for VL and EID.

Facilities requesting a VL test currently complete a standard MOH Viral Load Lab Requisition Form (Annex II). Data on this form include:

- Facility name and email address,
- Master Facility (MFL) code,
- Health care provider number,
- Patient identification number (CCC number),
- Date of specimen collection,
- Reason for test (e.g., routine VL, treatment failure, etc.), and
- Clinical details of the patient.

These data and the VL results are entered into the regional laboratory information system and further uploaded into the NASCOP National VL Database. CBS data can be derived from either of these sources.

Documentation of CD4 tests and results, unlike VL, remain at the local level; that is, either a local laboratory system or lab logbook. Given that there is no national or regional repository for CD4 results, such data will be accessed from clinical records. EID data are also limited in that current testing request forms/documents sent to regional laboratories only include the child's HIV-Exposed Infant (HEI) identifier. At present, results cannot be linked from the EID database to a child or mother who has been reported in the CBS database. This limitation however should be taken into consideration as the surveillance system matures, as EID results provide the diagnosis for a child and thus should be linked or reported directly to the CBS database.

4.4 Pharmacies

Both electronic and paper-based pharmacy information systems (PhIS) are used to track dispensing of ART. Mechanisms for the possible integration of PhIS data should be assessed, including variables available/required for CBS, and unique identifiers used. The key variables required in the pharmacy database for prescription data are, however, available through the Green Card, making PhIS of secondary importance. If there are facilities which use a PhIS but do not have an EMR, this may be an important alternative source of data.

4.5 Civil Registration and Vital Statistics Registries

Civil registration and vital statistics (CRVS) records are perhaps the most challenging to collect for the CBS system, at least currently. While the Green Card/MOH 257 collects death, cause of death (COD) is not specified, specifically whether a death was HIV-related or not. In addition, it has been observed that death is generally not well-documented either in paper or EMR records. Vital statistics registries (outside health facilities) that collect data for all deaths as well COD may provide a secondary, but powerful, source of data. At present, vital registration data are not available in an electronic format, are generally filed by date of death (with different formats for health facility verses community deaths) and do not involve unique identifiers other than names of the deceased and date of birth. While there are a number of logistical hurdles, there is ongoing work to improve the CRVS program. Integration with CRVS is part of the longer-term strategy for CBS. In the short term, routine program evaluation of patients who are lost-to-follow-up (LTFU) to assess proportions who have died as well as data from mortuary surveys may be used as sources of data on deaths among HIV patients as well as HIV-related deaths.

Table 7: Overview of Data Sources for CBS

Facility/data source type	Examples of service points	Data Tool	Sentinel event(s) and associated data elements
HTS Sites	HTS at the outpatient and in-patient departments: • Provider-Initiated Counselling and Testing (PITC) • Voluntary Counselling and Testing (VCT) • Antenatal Clinics (ANC) and family planning (FP) clinics • Voluntary medical male circumcision (VMMC) • Sexually transmitted infection (STI) and TB clinics • Targeted community/home-based testing services • Drop in centres (DICE) serving key populations	• MOH 362 HTS register • Mobile HTS applications • EMR HTS module • Linkage register	• Date of HIV diagnosis
HIV Care and Treatment Facilities	• Dispensaries • Health centres • Sub-county and county hospitals • National and referral hospitals • PMTCT • Drop in Centres(DICEs)	• EMR • MOH 257 (Green Card) • Linkage register	• Date of diagnosis • Date of enrolment • Date of ART initiation • Date and result of CD4 and VL test • Date of death
Laboratories	• Facility-based laboratories • Regional laboratories • National reference laboratories	• National VL Database • MOH 362 • EID Report form	• HIV diagnosis • PCR for EID • CD4 results • VL results
Pharmacies	• Facility-based pharmacy	• Web Antiretroviral Dispensing Tool (ADT)	• Change in ART regimens • Date of ART initiation
Vital Statistics Registries	• Vital statistics registries • Mortuaries	• Registry of deaths	• Death • Date of death • Cause of death

4.6 Reporting Scenarios

Data capture will be done daily either at the point of service provision or retrospectively by a designated CBS focal person. CBS implementation SOPs provide detailed guidance on reporting procedures from the different types of record documentation and storage system at the facility level. In summary, Table 8 shows the four main scenarios found at facilities and the overall action that should be taken by the facilities for CBS implementation.

Table 8: Summary of Reporting Scenarios

Facility Scenario	Action
EMR and eMobile HTS	Since the eMobile HTS links EMR, this is the ideal scenario. Data will be pulled from EMR into national data warehouse (NDWH). From the NDWH, cases are sent into the CBS database.
EMR + paper HTS	Data should be captured in the MOH 362 paper register and then entered into the HTS module within the EMR.
eMobile HTS only (no EMR)	The data collected in the eMobile HTS are uploaded to eHTS or intermediary server. From here, all HIV-infected cases are sent to the CBS database. Capture of sentinel events beyond HIV diagnosis is not required.
All paper	At minimum, where possible facilities would need to investment in using an eMobile HTS applications or another electronic data capture tool. Simultaneously, the MOH is working on alternative solutions for paper sites.

5. The HIV Case-Based Surveillance Data Management System

The design of the HIV case-based surveillance data management system shall follow relevant WHO guidelines on case surveillance but respond appropriately to Kenya's country-specific context. The data management system will comprise a central database and related applications for managing all reported cases of HIV. The HIV CBS database management system will be able to receive and store cases from various data collection systems. The data management system will provide features to clean, de-duplicate, transfer and analyse the data. These guidelines do not specify how system requirements should be met but rather provides guidance on the minimum requirements of such a system.

The data management system will leverage existing systems, be integrated with other case-based disease surveillance systems or be stand-alone specifically designed for HIV CBS. Considerations while developing a CBS data management system are listed in Section 5.1 below.

5.1 Database Considerations

Leveraging existing information systems

Kenya has invested substantial resources in program monitoring and surveillance data systems; in particular, electronic systems. To implement CBS at scale, it would be advantageous to leverage these information systems where applicable and automate the data extraction process as much as possible. The information system support for HIV CBS should:

1) Maximise the utilization of the existing EMR and related systems in health care facilities
2) Anticipate that eMobile HTS application will increasingly be deployed in facilities that offer HTS and community-based testing programmes
3) Incorporate information flow from laboratories and pharmacies
4) Incorporate paper-based data from HTS and health care facilities through data entry into an electronic data collection process that enables transmission of data to the CBS system.

In situations where additional information system components will be developed or established, the following principles should be followed:

- Adopting an existing health information system, components or platform is preferred over new software development.
- Achieving cross-functionality between existing systems and CBS will require understanding and addressing any barriers to integration. .
- The country has established eHealth strategy and policy documents; therefore, additional components should follow guidance from these documents.
- There should be deliberate efforts to align information system needs specific to HIV programmes with Kenya-wide health information infrastructure.

Levels of data management

Under the Kenyan constitution, health is a function that is devolved to the county level, thus the CBS system shall support multiple levels of data management, specifically health facility, sub-county, National and County. Nonetheless, in order to provide a national picture of the HIV epidemic, and support standardised reporting, de-duplication and transfer of cases across administrative units, the data management system shall have a central repository. Existing systems such as KHIS, VL, EID, and the national TB programme electronic reporting system Treatment Information from Basic Unit (TIBU) can provide models for systems based on centralized database with decentralized controls. The system will need to implement role-based access and controls that define which features and data are available to users at different administrative levels within the system.

Data extraction

Much of the data required for CBS are already collected in paper-based registers, EMRs at health facilities and in central laboratories. Extraction of case report data from sites with EMRs may save time and reduce the need for additional human resources. There are several approaches available for extracting case reports and sentinel events from EMRs. One approach is where EMRs upload an updated listing of all cases and their sentinel events on a regular basis to a central repository. Another approach is to use a health information exchange (HIE), where each EMR or other data system sends only new cases and sentinel events since the last transmission.

The CBS system shall initially build upon the existing national data warehouse infrastructure, which implements the first approach. The details and benefits of each approach are detailed in Annex III. For data sources using paper registers, we will leverage existing initiatives to capture all HTS data in eHTS, supporting the updating/upgrading of such applications to fit CBS needs.

Record linkage and de-duplication

Linkage refers to linking records across disparate sources into a single case record. Linking laboratory, clinical and vital statistics reports for a single case is one example of record linkage. Linkage is crucial for the CBS system given the multiple data sources for sentinel events. De-duplication and linkage are separate, but related, concepts. De-duplication refers to the elimination of duplicate case reports that may be present in the system due to, for example, misspellings or changes in unique identifiers (15).

For the system to be considered reliable and of high quality, it is expected to have a minimal number of duplicate cases resulting from changes in information, for example name or residence. It is therefore critical for the surveillance system to differentiate individuals who had been previously reported into the system from truly new or distinct cases. Without the ability to make this distinction, the surveillance system will over-estimate the true number of cases and misclassify outcomes. De-duplication is easier closer to the source of the case where contextual information is often available to help identify and resolve the duplicates. Therefore, the system should provide features to support de-duplication of patients starting from the lowest level possible.

The CBS system will make the de-duplication closer to the source system through a duplicates removal feature available in the Data Warehouse API client tool (DWAPI). This tool shall be running regularly against the source EMR server and will check for possible duplicates at the facility level. The tool shall provide the health worker with a report of these cases for further clerical review and merging using the patient merge features within the EMR. De-duplication is also required above the primary point of care due to lack of integration of systems across facilities and programmes. This underscores the need for a centralized data management system to facilitate the de-duplication of patient records being reported to the CBS system. The system must track and flag suspected and confirmed duplicates so that they are not reported as new cases. The process for de-duplication shall be standardised, documented in an SOP and conducted on a routine basis.

5.2 Ensuring Data Quality

There are six core data quality dimensions; completeness, uniqueness, timeliness, validity, accuracy and consistency, which should be considered at data collection points. Measurement of these elements is further elaborated in chapter 8. Continuous monitoring of data should be enhanced, for example at the health facility EMR. Within the CBS system, data quality checks will be applied at all levels. At the facility level, EMRs will have inbuilt data entry controls while the application developed for paper sites shall include checks to enhance completeness and accuracy. Data cleaning is generally most effective where data are generated. Furthermore, cleaning data at the facility level may improve patient-level clinical records and ultimately quality of patient care.

At the sub-county and county levels, data quality monitoring shall involve supervision exercises, checks on completeness and accuracy of reports and resolution of duplicate case reports. Triangulation between reports from KHIS and CBS shall be used to identify gaps in reporting and other data quality issues. Data quality audits (DQA) shall be conducted at facility level as per the guidelines for data quality assessments for the HIV program in Kenya. Data quality will also be one of the aspects reviewed when conducting periodic evaluations of the surveillance system's ability to serve its purpose (Section 8).

5.3 Data Analysis and Reporting

Surveillance units and programmes require standardised reports to document the status of the epidemic. There is often a need to address the requirements of different users, for example by level (national, county, sub-county) or purpose (programmatic or surveillance). Thus, the system must provide flexibility and visual display (such as dashboards) needed by different users for decision making and programming. Chapter 7 describes data analysis and dissemination in more detail.

6. Data Security and Confidentiality

6.1 Confidentiality of HIV-related Data

The HIV & AIDS Prevention and Control Act of 2006 (revised in 2012) states that HIV-related data on individuals must be treated as confidential (Part 5, Sec 20-22) and provides the authority to the MOH to specify guidelines or regulation regarding the use of identifiers with respect to HIV-related data. The Data Protection Act, 2019 also focuses on confidentiality (11). Data storage and handling shall comply with the applicable laws and regulations.

6.2 Data Security

The CBS system will carry patient identifiable information (PII) alongside sensitive clinical information regarding HIV diagnosis and treatment. These data will be captured both electronically and/or on paper depending on infrastructure at the facility or point of capture. These data will eventually be transmitted electronically into a centralized national surveillance database. As such, MOH should select and routinely evaluate technologies and security policies and procedures that ensure data security for the CBS system. Acceptability of the system relies on its ability to gather, store, use and transmit information in a secure manner. All programme processes and procedures should be implemented to ensure they protect patient confidentiality and promote data security. In particular, data from sites providing services to KP are sensitive considering existence of laws and policies prohibiting sex work, drug use and same-sex relations.

The overall architecture of the CBS system shall provide both a logical and physical separation between the patient PIIs and clinical data. This architecture is further illustrated in Annex IV.

A data security sub-committee or other body shall be put in place to develop, review and advise the surveillance team on surveillance information security including development of relevant SOPs. For electronic data, the data security sub-committee will establish the relevant security standards as per the guidance of the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC) 27002 information security standards specifications. The CBS servers and databases should be hosted at locations that assure the following three aspects of security are attainable:

Physical and Environmental Security:

- Data must be stored in safe physical environments.
- Physical access to premises and support infrastructure must be monitored and controlled.
- List of personnel authorised to access the CBS server is reviewed and updated periodically (at least bi-annually).
- The date and time of entry and departure of personnel accessing the CBS servers along with the reasons for access must be recorded in a register maintained and controlled by site security or reception.

- Both paper and electronic data must have process and procedures for their data security.
- Off-site or and back-up storage systems must also be secure with appropriate sanctions in place should security be breached in accordance with applicable laws.

Human Resource Security:

- All personnel accessing the CBS infrastructure must formally accept a binding confidentiality or non-disclosure agreement concerning personal and proprietary information provided to or generated by them in the course of interaction with the CBS system, with sanctions according to the laws in case of a breach.
- Staff working with surveillance data shall be trained in procedures for handling confidential data securely.
- The database administrator(s) shall be notified when authorised personnel are transferred, resigned, suspended, or released on long-term leave, or their employment is terminated for appropriate actions and/or access is revoked.
- Only authorized staff should have access to the data needed to conduct their job.
- Access to individual-level data shall be logged by the system.

Access Controls:

- User access to the CBS applications and information must be controlled in accordance to the data security and confidentiality SOPs, according to each user's role.
- Generic or test IDs must NOT be created or enabled on the CBS production system unless specifically authorized by the data security sub-committee.
- After a predefined number of unsuccessful logon attempts, security log entries and security alerts must be generated and user accounts must be locked out as defined by CBS security SOPs.
- The CBS system must require passwords to have a minimum length and level of complexity as defined in the SOPs.
- Passwords must not be written down or stored in readable format both externally or internally in the CBS database.
- Privileged access rights typically required to administer, configure, manage, secure and monitor the CBS system must be reviewed periodically by data security sub-committee and cross-checked by the system administrator(s).

Additional Security Measures Outside the Database:

- Utilization of encryption technologies to secure the data packets during transmission is required.
- Each participating system must be evaluated and authorised by the database administrator before inclusion into the CBS as source of data.
- The system should be certified, approved and issued with a key for data encryption for the purpose of transmission.

- Any patient identifiers stored in the system will be encrypted and only decrypted at the point of matching, linking and de-duplication.
- The CBS database shall be backed up periodically. The CBS data security sub-committee will define the threshold of time and method for backups. The backups should be stored on a separate location and used to recover the CBS database if needed.

6.3 Data Ownership and Sharing

Data collected through the CBS system shall be owned by the Kenyan MOH. These data shall be available to users for analysis and dissemination, with permission from the MOH through NASCOP. Policies or regulations governing how the CBS data would be shared within, across and outside of NASCOP and under what circumstances CBS data would be shared should be clearly described (14), such as a memorandum of understanding (MOU). These policies/regulations could include direction on how surveillance data will be used and disseminated as well as how and when data can be released to a third party, for instance a multi-national health agency like WHO, UNAIDS. The policies/regulations should require that a data sharing agreement with these third parties is signed before sharing and ensure that no protected information (identifiable information) is released to the third parties (16).

7. Data Analysis and Dissemination

7.1 Analysis and Interpretation

The analysis and interpretation of surveillance data should be done on a routine basis by both surveillance and programme managers. This facilitates various perspectives of data interpretation by those most familiar with the data, how they were collected, their weaknesses and their limitations. Inclusion of different decision-makers also facilitates the access and use of case-based reporting data for programme planning and response.

In analysis, data from the HIV surveillance system can be looked at independently but can also be complemented by programmatic, research, census, STI and survey data to better understand the epidemic. Prior to analysis, quality of the data and their limitations must be known and documented. Data quality assessment should examine the ability of the data to answer the key questions of interest (e.g. prevalence, risk factors, demographics of HIV positive, distribution of cases by geography). To do so, several aspects of the data should be examined to determine potential limitations of the analyses. These include, but are not limited to, completeness, accuracy, missingness of key variables, de-duplication of cases, and timeliness. Analysis should describe the characteristics of those diagnosed with HIV, where they are being diagnosed, the proportion receiving treatment, clinical indicators, mortality and trends in disease progression over time. Annex V provides sample tabulations.

7.2 Dissemination

Data derived and disseminated from the surveillance system should take into consideration the needs of primary users. There should be consensus on the standardised outputs resulting from the database including the different stratifications (age groups, sex, rural, urban, marital status, etc.). The standardised dissemination material needs to be presented in the same format across the different levels and over similar time periods to accommodate trend analysis. Examples of some of the analysis that should be presented in the dashboards are included in Annex V. Additionally, the dashboards and reports should be flexible to allow for individualised analysis of the data based on the interest of the user.

To encourage the use of data from the surveillance system, the results from analyses must be disseminated at various levels (national, county and sub-county). There are various dissemination avenues for surveillance data. These include:

- *Annual surveillance reports:* These reports contain descriptive statistics, trends in the epidemic, risk factors with disaggregation by variables of interest including age, sex, areas of residence (rural/urban) and trends in these indicators over time. See Annex V for examples of analyses that can be done for a surveillance report.

- *Strategic reports:* These reports contain a summary of the state of the HIV infection. Data captured in these reports provide information on HIV surveillance activities and efforts to inform the response to the epidemic.

- *Data sheets:* These are brief descriptions on specific areas or topics of interest, for example, a look at the HIV epidemic among female sex workers.
- *Presentations:* These are visual presentations of information covered annually in all of the other formats of data dissemination to facilitate quick and effective data consumption by the general public, decision makers, public health scientists and community organizers.

8. Evaluation of the Surveillance System

Periodic evaluation of the HIV CBS system will ensure a relevant and responsive system that can meet its objectives. An evaluation of the CBS system will be conducted according to a pre-determined schedule, for example, annually. The governance mechanism established for the CBS system will also be responsible for ensuring that evaluation of the system is conducted as scheduled and to standard. A crucial initial step in evaluating the system is engagement with stakeholders (the owners, operators, and customers of the system) to ensure that the planned evaluation addresses appropriate questions, targets the pertinent attributes of the surveillance system and yields acceptable findings. Besides the attributes of quality, other elements to evaluate include processes for records matching and deduplication to ensure uniqueness of records, the economic features and the usefulness (how the data and surveillance indicators are used in decision making) of the surveillance system. The quality of the surveillance system will be defined by attributes that are well described in the updated guidelines for evaluating public health surveillance systems produced by CDC (17). These include:

Simplicity

This refers to both the structure and the ease of operation of the surveillance system. A simple way to assess this could be drawing a chart describing the flow of data and the lines of responses in the surveillance system. To ensure simplicity of the CBS system, the case-based surveillance form will be made easy to read and understand, and will include only the elements needed for the CBS. Making the HIV case definition simple will also contribute to simplicity of the CBS system.

Measures that could be considered in assessing simplicity may include:

- The amount and type of data required to ascertain a case
- The amount and type of data being collected for each case
- The amount of time taken in data collection, transferring, entering, editing, storing, and back-up
- Staff training requirements (job aids, SOPs, etc)

Timeliness

Although not as critical as in other infectious diseases that require immediate notification, this attribute is still important in a CBS system for HIV. It reflects the delay or speed between steps in the surveillance system. For an HIV CBS system, the steps may be:

1. Detection or diagnosis of sentinel events
2. Data entry (depending on set-up, e.g. paper-based case-based surveillance form may need to be entered)
3. Reporting of the event to the CBS system
4. Data analysis and interpretation
5. Response/ intervention activities undertaken and feedback to stakeholders.

For instance, the time taken between step 1 to 2 of the CBS system can be measured e.g. time between detection of an HIV case and reporting this to the surveillance system. The acceptable time interval between steps depends on availability of options for prevention and control measures that can be taken. It is also important to determine timeliness of trend identification or detection of effect of control measures taken.

Acceptability

Organisations involved in the CBS system are required to participate willingly. Quantitative measures that could be used to asses this attribute of the CBS include: facility reporting rate, completeness of report forms and timeliness of data reporting. Qualitative, in-depth interviews with stakeholders should be conducted to explore the level of willingness of providers and facilities to participate in the surveillance system. It is worth noting that other attributes of the CBS system may influence acceptability of the system. These include ease of using tools such as electronic systems (simplicity), reliability of the electronic system being perceived as part of routine work.

Flexibility

Flexibility reflects the ability of the HIV CBS system to adapt to changing information needs or operational conditions with little additional time, personnel or funds. A flexible CBS system can accommodate variations in reporting sources, new sentinel events, changes in HIV case definitions or changes in technology used. A suggested approach to assess this attribute is to conduct a retrospective observation of how the system has responded to changes. If no efforts have been made to adapt the surveillance system in any way, assessment of flexibility may be difficult. It is also generally thought that simpler systems are likely to be more flexible.

Data quality

The completeness and accuracy or validity of the data recorded in the surveillance system reflect the data quality. Various approaches can be employed to ensure quality of data collected. These include consistency checks, completeness checks, triangulation of responses from different sources and data cleaning queries. One approach to assess this could be examining the proportion of "unknown" or "blank" responses to items on case report forms. For more comprehensive evaluation of completeness and validity of data in the CBS system, special studies may need to be designed, for instance, comparing data values in the CBS with "true" values obtained through patient interviews or reviews of sampled patient records.

Table 9: Elements of data quality and operational definitions

Data Quality Element	Operational Definition
Accuracy	Accuracy refers to the extent to which data reflect the actual/correct information. *Example:* If there are 3000 confirmed newly diagnosed HIV cases entered into the system then the CBS system should reflect the same number.
Completeness	Completeness reflects the inclusiveness of data such that data that all expected variables are captured correctly. *Example:* All the elements of an individual are captured as per the instructions of the tool.
Reliability	Reliable CBS data are those that are complete, accurate and useful for calculation of intended indicator and are not subject to changes over time and in subsequent analysis. *Example:* The number of patients enrolled in ART as at a particular date should be consistently the same. Data are assumed to have been cleaned before reporting.
Precision	This means that the data measures the same indicator in the same way every time. *Example*: Alteration of measurement of age as at last birthday should be consistently recorded as such. Any other interpretation such as "age attained this year" would result in a different age. Additionally systems should have a reference date for age e.g. 15th June where the actual date is unknown but the year of birth is provided.
Timeliness	Data are timely when they are up-to-date (current), and when the information is available on time. *Example:* Data should be entered by the 15th of every month nationally. Any data that re not entered by this date are untimely. Reporting timeliness has to do with by when CBS data are uploaded to the NDW.
Integrity	Data have integrity when they are not spurious and have not been deliberately manipulated. *Example:* Manipulation: Reporting numbers where no actual service has been provided to meet targets. Falsely altered: the data clerk enters fabricated data.
Confidentiality	Confidentiality means that clients and patients data remain private and are not disclosed to unauthorised persons. Laws and regulations for data security are adhered to with penalties imposed. *Example:* Client files are kept in locked cabinets and/or in password protected files. Details for maintaining of electronic records are provided in section 5.

Sensitivity

At the level of case reporting, the sensitivity of the surveillance system refers to the proportion of diagnosed cases of HIV that is detected by the CBS system. At the data analysis and use levels, sensitivity is the ability of the CBS system to monitor and detect changes in the number of cases over time. The primary emphasis of assessing sensitivity is on estimating the proportion of all diagnosed HIV cases in the population under surveillance (usually an estimate) that is being detected by the HIV CBS system. This assumes that most reported cases are correctly classified. Sensitivity can be calculated for each data source (to determine effect of eliminating a data source from surveillance system), a combination of data sources or for each year of surveillance. Of note, is that a surveillance system without high sensitivity may still be useful in detecting trends as long as the sensitivity level remains reasonably constant over time.

Representativeness

To allow generalization of findings from a CBS surveillance system to the larger HIV infected population, the data from the surveillance system should accurately reflect the characteristics of the PLHIV under surveillance. The CBS system is representative if it accurately describes the occurrence of the sentinel events over time and their distribution in the population by place and person. Approaches to examining representativeness include designing special studies that compare CBS data with external population-level data sources. For instance, a study can identify a sample of cases from the KHIS or national KAIS, DHS or KENPHIA data and examine statistical measures of population variables then compare these measures with measures of the same population variables among CBS cases. One important outcome of evaluating representativeness is the identification of a population subgroup that may be systematically excluded from the reporting system which should lead to corrective actions.

Stability

Stability refers to the reliability and availability of the surveillance system, i.e. its ability to be operational when needed. Methods that can be used to measure the stability of the CBS system include:

- Percentage of time that the system is fully operational
- Number of unscheduled outages and down times for the system's computer
- The desired and actual amount of time required for the system to collect or receive data. These measures of stability could be assessed against the purpose and objectives of the surveillance system

9. References

1. National AIDS Control Council (NACC), National AIDS and STI Control Programme (NASCOP) K. Kenya HIV Estimates Report 2018. October, 2018 [cited 2017 Mar 16];1–28. Available from: https://nacc.or.ke/wp-content/uploads/2018/11/HIV-estimates-report-Kenya-20182.pdf.

2. National AIDS Control Council (NACC). Kenya HIV prevention response and modes of transmission (MOT) analysis. [Internet]. 2009. Available from: https://nacc.or.ke/images/stories/Documents/KenyaMOT22March09Final.pdf

3. National AIDS and STI Control Programme (NASCOP). Kenya. Integrated Biological and Behavioural Surveillance Survey among Key Populations in Nairobi and Kisumu. 2011; Available from: https://www.nascop.or.ke/?p=105

4. Joint United Nations Programme on HIV/AIDS (UNAIDS). 95-95-95 Understanding Fast-Track. Accelerating Action To End The AIDS Epidemic By 2030. https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf

5. National AIDS and STI Control Programme, Ministry of Health K. Case based Surveillance of HIV in Kenya: Results of a pilot conducted in Kisumu and Siaya. 2017;

6. UNAIDS/WHO Working Group. Guidelines for Second Generation HIV Surveillance : an update : know your epidemic. WHO Guidelines [Internet]. 2011 [cited 2017 Apr 19]; Available from: www.who.int/about/licensing/

7. World Health Organization. Consolidated strategic information guidelines for HIV in the health sector [Internet]. 2015 [cited 2017 Apr 19]. Available from: http://apps.who.int/iris/bitstream/10665/164716/1/9789241508759_eng.pdf?ua=1&ua=1

8. World Health Organisation. Consolidated guidelines on person-centred HIV patient monitoring and case surveillance. 2017.

9. Ministry of Health K. Standard_Case_Definitions_for_Priority_Diseases_in_Kenya-_Integ.

10. Data Protection Act, 2019. http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/2019/TheDataProtectionAct__No24of2019.pdf

11. Lecher S, Williams J, Fonjungo PN, Kim AA, Ellenberger D, Zhang G, et al. Progress with Scale-Up of HIV Viral Load Monitoring — Seven Sub-Saharan African Countries, January 2015–June 2016. MMWR Morbidity Mortality Weekly Report [Internet]. 2016;65(47):1332–5. Available from: http://www.cdc.gov/mmwr/volumes/65/wr/mm6547a2.htm.

12. Civil Registration Services (CRS). Kenya Vital Statistics Report 2014. Nairobi, Kenya: Civil Registration Services, CRS; Statistics Division. Government of Kenya; 2015.

13. UNAIDS. Considerations and Guidance for Countries Adopting National Health Identifiers. 2014;1–112. Available from: http://www.unaids.org/sites/default/files/media_asset/JC2640_nationalhealthidentifiers_en.pdf

14. Waruru A, Natukunda A, Nyagah L, Kellogg TA, Zielinski-Gutierrez E, Waruiru W, Masamaro K, Harklerode R, Odhiambo J, Manders E, Young P. Where No Universal Health Care Identifier Exists: Comparison and Determination of the Utility of Score-Based Persons Matching Algorithms Using Demographic Data. JMIR Public Health Surveillance 2018;4(4):e10436) doi:10.2196/10436

15. National Center for Immunization and Respiratory Disease (NCIRD). 2013. Immunization information systems patient-level de-duplication best practices URL: https://www.cdc.gov/vaccines/programs/iis/interop-proj/downloads/de-duplication.pdf

16. NASTAD. System Governance and Partner Engagement_AC_NASTAD.

17. Centers for Disease Control and Prevention. Updated Guidelines for Evaluating Public Health Surveillance Systems Recommendations from the Guidelines Working Group. MMWR Recommendation Reports / Centers Disease Control [Internet]. 2001 [cited 2017 Apr 19];50(RR-13):1-35; quiz CE1-7. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm

18. Broyles D, Crichton R, Jolliffe B, Sæbø JI, Dixon BE. Chapter 10 - Shared Longitudinal Health Records for Clinical and Population Health. In: Health Information Exchange: Navigating and Managing a Network of Health Information Systems. Academic Press; 2016. p. 149–62.

Annex I: HIV Case-Based Surveillance Form

Version: 25 Sept 2018

A. REGISTRATION

Person Completing Form: *First Name* _____ *Last Name:* _____

Reporting Date: __ __/__ __/__ __ __ __
Reporting County: _____
Reporting Sub-County: _____
Reporting Facility Name: _____
Reporting Facility MFL code: |__|__|__|__|__|__|

Report type:
☐New case
☐Update of existing case (new sentinel event)
☐Correction of previous case report
☐ Transfer In

B. PATIENT PROFILE / DEMOGRAPHICS

Patient Name: First: Middle: Last:

Date of Birth (DD/MM/YYYY): __ __/__ __/__ __ __ __

Age as documented on MOH257: |__|__|

Age as documented on HTS register: |__|__|

Phone No.: |_|_|_|_|_|_|_|_|_|_| ☐None
Ward: _____
Village/Estate: _____
Sub-County: _____
County: _____

Sex: ☐Male ☐Female

Population type: ☐Gen pop ☐Key Pop **If Key Pop, tick one** ☐MSM ☐FSW ☐PWID

CCC No.: |__|__|__|__|__|-|__|__|__|__|__| ☐ None

National ID No.: |_|_|_|_|_|_|_|_|_|_|_| ☐ None
Other unique No[1]: _____

If patient is a child, then fill below if known for additional profile information to assist in patient identification.

HEI number: |_|_|_|_|_|_|_|_|_|_|_| ☐ None **Mother's CCC number:** |__|__|__|__|__|-|__|__|__|__|__|

Mother's Name: First: Middle: Last:

C. HIV DIAGNOSIS AND LINKAGE TO CARE/ANTIRETROVIRAL TREATMENT (ART)

Date of HIV Diagnosis: __ __/__ __/__ __ __ __ **Diagnosis by:** ☐ PCR ☐ Rapid Antibody test

Recency HIV infection test conducted?[2] ☐ Yes ☐ No **If yes, HIV infection recent?**[3] ☐ Yes ☐ No

Referred from (Place of 1st diagnosis)

☐ HBTC
☐ VCT
☐ OPD
☐ MCH
☐ TB Clinic
☐ IPD-Child
☐ IPD-Adult
☐ CCC
☐ Self-test
☐ Other (e.g. STI) _____

Client tested through Partner Notification Services (PNS)?[4]
☐ Yes ☐ No

WHO stage at enrollment to care: |__|
TB infected at enrollment: ☐ Yes ☐No
Pregnant at enrollment: ☐ Yes ☐No
Breastfeeding at enrollment: ☐ Yes ☐No

History of ART at enrollment:
☐ PrEP ☐ PEP
☐ PMTCT ☐None

Date enrolled in HIV care:
__ __/__ __/__ __ __ __
Date of ART initiation:
__ __/__ __/__ __ __ __

D. CLINICAL EVENTS

ART Regimen Changes
Date changed to 2nd line ART __ __/__ __/__ __ __ __
Date changed to 3rd line ART __ __/__ __/__ __ __ __

TB treatment[5, 6]
Date of start of TB treatment __ __/__ __/__ __ __ __

E. LABORATORY INVESTIGATION: VIRAL LOAD, CD4

CD4 sample collection date: __ __/__ __/__ __ __ __
CD4 test results: _____ Cells/μL **OR** CD4% _____

VL sample collection date[6] __ __/__ __/__ __ __ __
VL result _____c/ml ☐ <LLD [lower limit is __ __ c/ml]

F. STATUS

LTFU[7]: ☐ **LTFU as of Date:** __ __/__ __/__ __ __ __
Transfer out: ☐ **Transfer out Date:** __ __/__ __/__ __ __ __
Death: ☐ **Date of death**: __ __/__ __/__ __ __ __

G. DATA SOURCE USED (Select all applicable)

☐ HTS Register
☐ Treatment Preparation Register
☐ ART Register
☐ Individual Clinic Record (e.g. MOH257)
☐ ANC register
☐ HEI register
☐ Viral Load Register
☐ TB Register
☐ Other: _____

H. NOTES

1. **Other Unique Number**: Pull down with ANC number, DSS number and any other agreed upon
2. **Recency HIV test conducted:** This is a place holder
3. **HIV infection recent:** This is a place holder
4. **Client tested through PNS:** This is a place holder
5. **Date of start of TB treatment**: included as proxy for date of TB diagnosis
6. TB occurrences and VL results are **repeated events**. Report all occurrences.
7. **LTFU**: Not appearing for >90 days since last clinic appointment.

Annex II: Viral Load Laboratory Request Form

Version- June 2016

Ministry of Health

Viral Load Requisition Form

Facility Name …….. Date Sample Dispatched…… Time of sample dispatch……..
Facility MFL code ……..... VL focal person's name……..... Requesting clinician Name……
County Name ……..... VL focal person's phone number…….. Requesting clinician number……
Sub county Name ……..... Facility email……..... Clinician email……

Serial No	Patient Name	CCC No	DOB *(mm/dd/yyyy)*	Sex	If female, select the following *1 = Pregnant 2 = Lactating 3 = None of the above*	Sample type *(select from code below)*	Date of collection	Time of sample collection	Date started on ART	Current ART Regimen *(select from code below)*	Date initiated on current regimen	Indicate 1st line (1) or 2nd line(2)	Justification code *(select from code below)*	Rejection *(for reason select from code below)*
Y/N	*Reason*													
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														

Code for Sample Type: 1= Frozen plasma 2= Venous blood (EDTA) 3= DBS capillary (infants) 4= DBS venous 5= PPT Code for Justification: 1= Routine VL 2= Confirmation of treatment failure (repeat VL at 3M) 3= Clinical failure 4= Immunological failure 5= Single drug substitution 6= Baseline VL

Code for Sample Rejection 1=Improper collection 2=Incorrect container 3=Missing patient ID 4=Sample request & sample mismatch 5=Delayed delivery 6=Serum ring 7=Expired filter paper/tubes 8=Specimen processing delay 9=No requisition form 10=Improper packaging 11= Improper drying/shipment 12=Sample for patient > 18M 13=insufficient volume 14= poor collected DBS 15= Other (please specify)

ART REGIMEN CODES: 1= TDF+3TC (FTC)+EFV | 2= TDF+3TC (FTC)+NVP | 3= TDF+3TC (FTC)+DTG | 4= AZT+3TC (FTC)+NVP | 5= AZT+3TC+EFV | 6= ABC+3TC+NVP | 7= ABC+3TC+EFV | 8= ABC+3TC (FTC)+DTG | 9= AZT+3TC+NVP | 5= ABC+3TC+LPV/r | 6= AZT+3TC+LPV/r+ RTV | 7= TDF+3TC (FTC)+ATV/r | 8= TDF+3TC (FTC)+LPV/r | 9= AZT+3TC+LPV/r | 10= AZT+3TC+ATV/r | 11= AZT+3TC+DRV/r | 12=Other please specify

Annex III: Database Examples

Below are examples to potential database examples for Kenya, which need not to be mutually exclusive:

- A solution based on a **data warehouse** (DW) whereby data from EMRs, testing sites, laboratories, pharmacies, vital registries and other sources is collected centrally. Case reports and sentinel events are then abstracted by 'mining' through this data.

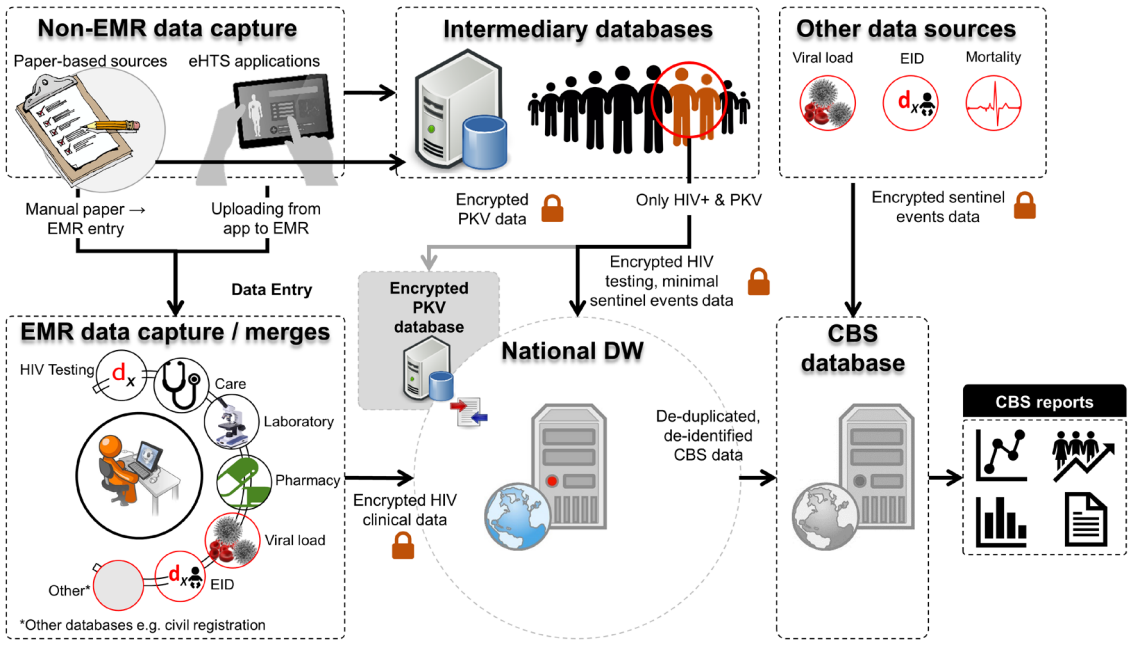
- A solution based on an **electronic** HIE. An HIE presents a network of linked information systems that supports the exchange of health data between information systems through a set of standardised and mutually agreed upon transactions.

Data Warehouse Solution:

In this scenario, a CBS database will be built off routine data collection systems including the proposed DW for EMRs to minimise new software development. As such, the majority of the issues that may arise with data within the CBS database will need to be resolved at the source, i.e., the EMRs (which include the green cards and case files), laboratory databases and other electronic data sources such as field testing data. Figure 1 below is an illustration of CBS database structure where several data sources would feed into the database for storage, cleaning, and analysis. The data analysis outputs would then be shared through a dashboard with standardised reports representing indicators of interest at different levels. Dashboards would also be an opportunity to export raw (de-identified) data for specialized analysis or study by authorized entities.

The CBS DW should be capable of receiving or extracting electronic data from EMRs, laboratory and pharmacy information systems, stand-alone HIV testing sites and other sources of data for HIV case-based reporting. Currently, the existing EMR data warehouse has great potential for contributing to the CBS data warehouse. The DW currently collects individual-level data on all HIV-infected individuals receiving treatment and care at facilitates with EMR systems in Kenya. The data collected in the DW are based on the MOH 257 (Green Card), which has most coverage of variables required for case-based reporting including diagnosis data, ART uptake, laboratory results, WHO stage and vital status. It is currently the only national level database with individual-level data on HIV-infected individuals.

Figure 3: A simplified CBS database structure



To improve on the existing DW for CBS, linkage of laboratory information systems (LIS) to the DW is advised. LIS would add a critical layer of information for both validation and coverage of lab-related information in the CBS database. Furthermore, LIS are a unique opportunity to capture HIV-infected individuals who receive services at paper-based sites and that document HTS data on MOH lab registers. Additionally, laboratory data repositories such as the VL database can serve as an interim source of data with minor modifications on identifiers collected (i.e., addition of names and CCC numbers to database) as the different laboratory systems are eventually linked to the CBS database.

Stand-alone HIV testing facilities should also receive a concerted effort in inclusion within the CBS database in the early stages of implementation of CBS. These are critical in identifying and reporting individuals who test HIV positive but do not access services, a critical element in documenting the continuum of care in the epidemic. PhIS could also support data needs for the CBS system especially if documentation of regimens and validation of status of individuals (e.g. alive and on treatment) is required.

As the country progresses to electronic systems in many of its institutions, it is important to prepare the CBS system to plug-in to such systems of data sources. These include vital statistics registries, which, when made electronic, can be routinely used to either update current status of reported cases or report new cases that had not been previously captured by the surveillance system. Furthermore, the CBS programme should identify, periodically update and maintain a comprehensive list of all reporting sites. Ideally all reporting facilities should have an individual responsibility for communicating with the surveillance programme and for ensuring compliance with reporting processes and regulations regardless of the manner in which data are collected.

Table 10: Data warehouse solution for CBS

Data Warehouse	
Advantages	**Disadvantages**
• Relatively easy to implement, as EMRs don't need to be adjusted – they just need to be able to export their data • Maximum flexibility in terms of data formats, provided that the data warehouse developers are agile and flexible, data can be provided to the DW in any type of format and then be absorbed through automated extract, transform and load programs • No complex information technology (IT) infrastructure and reliable continuous Internet activity required in the 'field' – only the national data warehouse needs reliable internet • Can work offline, by just transferring datasets on encrypted portable hard disks	• Requires a highly-skilled team with strong data warehousing skills to develop the data warehouse. The team should not be biased towards a single EMR, but willing to accept data in any of the formats used by the three most popular EMRs in the country • Data may not be fully comparable between various datasets, as the lack of strict standards may also mean that indicators do not always have the same meaning. For example, an *encounter* in one EMR may mean something different from a *visit* in another EMR. The DW team therefore needs to insist on data dictionaries • Some data may be incomplete because some EMRs may collect more data than others. A decision would need to be made to only collect the common variables or to collect everything and accept blanks where no data are available.

Health Information Exchange Solution:

The notion of exchanging health information is not new, but the concept of establishing a framework to connect the systems in a network in standardised ways is more recent. While originating from settings with strong health care systems and health information infrastructure, adoption in low- and middle-income country settings is also emerging (18). HIEs are intended to broadly solve the interoperability challenge by establishing policies and mechanisms for each system to adhere to.

The original motivation for HIEs is to coordinate information exchange in a complex distributed health services delivery setting. This scenario involves introducing the core pieces of an HIE that are essential to HIV CBS, allowing a process to be established where additional functionalities are added to the architecture by both adding additional components and then evolving the functional capabilities of components already introduced. The approach is forward looking in that it anticipates that other health domains will also be able to leverage these same core components of the HIE.

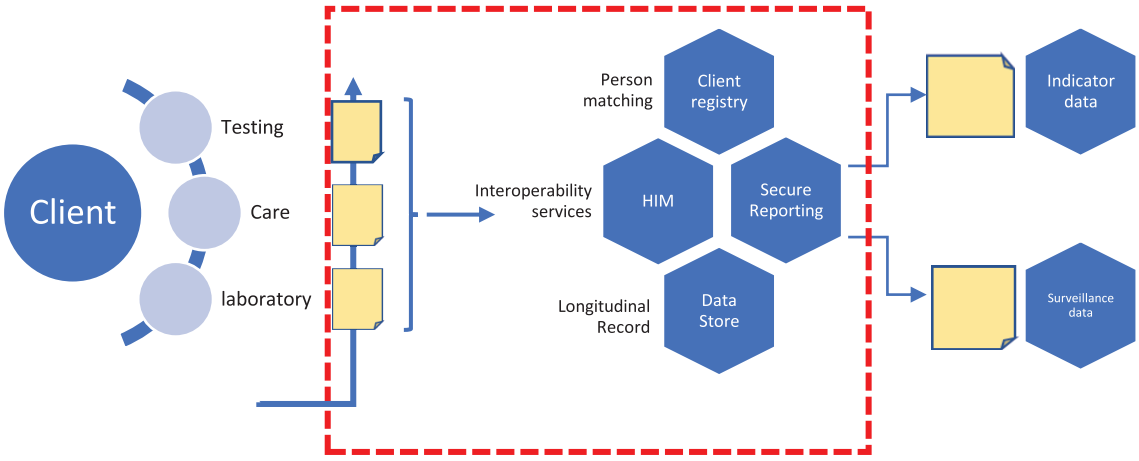
The core essential architectural components of an HIE are depicted in Figure 4 and are:

- Health interoperability layer, this service-oriented middle layer is referred to as a Health Information Mediator(HIM). The interoperability layer coordinates the sequence of exchanging information between different systems.

- Client registry. This component is also referred to as a Master Patient Index (MPI). The client registry provides a unified registry of patients known to any of the information systems that produce or consume data. The client registry will typically have mechanisms to look up patients and link entries that are found to belong to the same person.

- Clinical data repository. This repository holds health care data for each client known in the client registry. It is up to the individual connected systems to decide what information is exchanged.

Figure 4: CBS information system based on the architecture of an HIE.



The dashed box around the clinical documentation sources and the four HIE components represents the secure clinical data environment.

The figure also specifically shows a secure reporting module, integrated with the clinical data store. It is this component that will transmit data from the HIE to a surveillance data repository. In the process, the data will be de-identified. It is important to understand that the components in the red box are centrally hosted, which means that client identifiers are indeed stored at a central location.

An HIE relies heavily on development and subsequent confirmation to standards. What this means practically is that standards need to be developed that outline how data should be shared and that subsequently all EMRs and other systems that need to interoperate with the HIE will need to be updated to adhere to this standard. This may take a significant amount of time.

Table 11: Health information exchange solution

Health Information Exchange	
Advantages	**Disadvantages**
• The architectural foundation remains the same as the overall capabilities of the network of information system grows • Essential services that are associated with HIE frameworks are available and can be exploited, e.g. auditing of message transactions, advance security mechanisms • From a technological perspective, this is a 'cleaner' solution which can potentially lead to cleaner data and a full repository of electronic health records, irrespective of the disease area • A standards-compliant reference implementation for an HIE is available (Open Health Information Exchange) for a possible implementation • The Kenya national eHealth architecture proposes the adoption of an HIE as a national strategic priority	• Requires the adoption of interoperability standards which may take time as it is a stakeholder consultation process • Requires the implementation of these standards by every EMR and electronic system in use that does not already support those standards. If developers are not willing to do so, it may cause a challenge • Essential complexity around establishing an HIE is introduced at the start. An organisation that wishes to adopt an HIE approach must commit to developing competence in handling this complexity

A summary of these two options and their viability is described in Table 12 below.

Table 12: A comparison of the options and their viability.

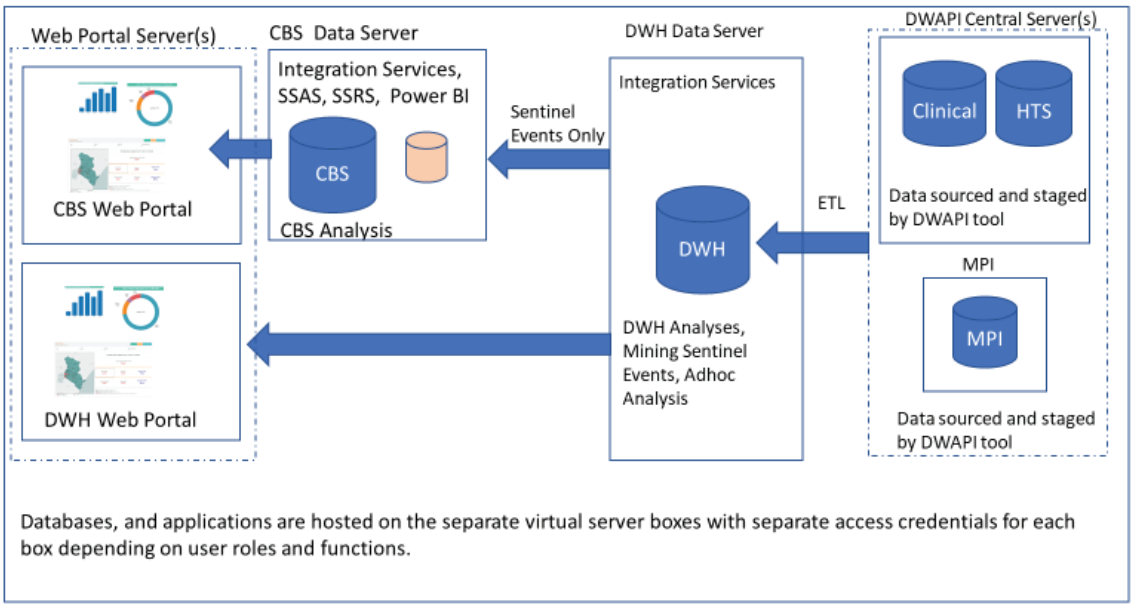
Area	DW	HIE
Concept	A system that integrates complete datasets obtained from EMRs and other systems throughout the country into a single database	An online electronic interface through which EMRs and other systems send small messages containing data for specific health related (sentinel) events and uses an online master patient index to identify existing cases
Dataset	• Contains transfer and storage of entire datasets, including all patient monitoring data • Includes multiple health domains (programme monitoring, CBS, possibly others) • Datasets stored in a central database	• Refined, granular health records, data transferred through individual messages • Focuses on a specific health domain (i.e. CBS) • Individual data elements stored in a central database
Data standards	• Does not necessarily require standards; a data dictionary describing the source data is sufficient • Data can be presented in multiple formats from various EMRs, and the DW team can write scripts and programs to import the data	• Requires the definition of rigorous and precise standards to be endorsed by the government; this may take a long time • The standards need to be developed before the HIE can be developed • All EMRs and other connecting systems must present the data through a predefined standard in order to interoperate with the HIE

Case de-duplication	• Can be achieved by attempting to match on patient identifiers and attributes in the data warehouse • Does not rely on a master patient index • Requires some identifiers to be stored centrally • Some de-duplication can be done by using encrypted identifiers	• Will be achieved all health data will be linked against the master patient index • Relies on an online master patient index, and fulltime online access to it (does not work offline) • All identifiers are stored centrally • Identifiers cannot be encrypted
Infrastructure	• Requires most investments at the top, i.e., where the DW will be housed • Lower databases can work offline	• Requires investments everywhere, and reliable and continuous internet connectivity is required • Offline use is not supported
Software development	• Medium complexity to develop • Crude, relatively fast to develop	• High level of complexity to develop • Refined, long development time
Impact on EMRs	• Does not necessarily require EMRs to be modified, they only need to be able to export data into a certain format	• Each information system connected to the HIE must support the interoperability standards established as part of the implementation of the HIE. • If standards evolve over time, maintenance of the implementations must follow the adopted changes in the standards.

Note on interoperability: For the CBS database to receive data from different sources, interoperability between the systems needs to be considered. Interoperability is the extent to which the different systems will exchange and interpret the data that are shared. For the CBS database, there needs to be an ability of the different IT systems and software applications to communicate, exchange data and use the CBS data for surveillance purposes. As such, development of data exchange schema and standards will be required to enable sharing of data across EMR, laboratories, pharmacies, HTS sites or any other sources of data for the CBS database. The level to which this needs to be done depends on the solution chosen. The DW requires less interoperability because some part of the data import will be handled by the data import team, and, as a result, standards can be developed "on the go". The HIE approach is, by definition, an interoperability approach and requires defining standards in advance before any data may be exchanged.

Annex IV: CBS System Architecture

Figure 5: NASCOP Data Centre Logical Infrastructure



The NASCOP data centre's underlying architecture incorporates virtualization, on-demand deployment, high availability, security and utilizes enterprise level open source software. Virtualization allows for a pool of resources to be shared based on application requirements for computing power, storage and network access demands. Software applications can therefore be deployed and scaled up or down without the need to procure additional physical resources.

Another key feature of having a virtualized architecture is the fact that it is possible to quickly add servers, without adding any physical hardware. Multiple servers are beneficial for several reasons:

- From a security point of view, it is good practice to store databases on servers that are not connected directly to the internet;
- It is good practice to conduct heavy data processing activities (for example, a cohort analysis or data mining) on a separate server, so that performance for other applications will not be affected ;
- Server administration, different access levels, privileges and rights can be applied independently to each server based on the guidelines and policies applicable to the functions of the said server and roles played by users or groups of users accessing the server; and,
- Any catastrophic event or disaster affecting one virtual server is not propagated on to the other servers as they are mutually exclusive.

The proposed architecture below shows that the performance of the different software systems can be significantly improved by having multiple distinct virtual servers each having a specific role. These three logical servers would all be provisioned at the NASCOP data centre on the same physical hardware as follows:

- **Data Staging Server:** This is where all the data from the EMRs and HTS apps is initially held. This server will primarily store the source data.
- **Master Patient Index (MPI) Server:** This is the server that receives and stores the PII's so that any patient data deduplication is done here;
- **Analysis Servers:** These servers conduct specific data analysis functions. These are separate virtual servers provisioned depending on the analysis role they support. For example, the DWH server supports Cohort Analysis, COP analysis and any other adhoc analysis required. The CBS server will conduct specific CBS analysis after mining sentinel events from the from the Staging Server and transforming this data into events that can then be reported on; and,
- **Web Application Servers:** These are the server(s) that will be connected directly to the internet for public access. Through these server(s), data from the Analysis Servers will be made available through visualization tools like Power BI and Tableau. For data security and patient privacy reasons, data is only accessible in a de-identified form.

Annex V: Sample Data Analysis Plan

Table A: New HIV diagnoses in [insert year] by age and sex

Age group	Sex		Total	
	Male	Female	N	%
0-9				
10-14				
15-19				
20-24				
25-29				
30-34				
35-39				
40-44				
45-49				
50-54				
55-59				
60-64				
65+				
Median age (yrs)				

Table B: Proportion of new HIV diagnoses who were started on ART in [insert year] by sex

Sex	New HIV cases	Number started on HAART	%
Male			
Female			
Total			

Table C: Patients who started treatment in [insert year]

Age group	Sex		Total	
	Male	Female	N	%
0-9				
10-14				
15-19				
20-24				
25-29				
30-34				
35-39				
40-44				
45-49				
50-54				
55-59				
60-64				
65+				
Median age at enrolment				
Median CD4 at registration (cells/mm^3)				

Table D: Number of newly registered patients at [insert # of clinics] by WHO staging at enrolment in [insert year]

WHO stage	Sex		Total	
	Male	Female	n	%
I				
II				
III				
IV				

Table E: Classification of newly registered patients at [insert number of clinics] by CD4 count at enrolment in [insert year]

CD4 categories	Sex		Total	
	Male	Female	n	%
< 200				
200 – 349				
350 – 499				
≥500				
Not known				

Table F: Classification of newly registered patients at [insert number of clinics] by VL suppression at [X time] in [insert year]

Viral load	Sex		Total	
	Male	Female	n	%
Not Suppressed (≥1000 copies/ml)				
Suppressed VL (<1000 copies/ml)				
Not known				

Table G: Immunological classification of newly registered patients at [insert number of clinics] by CD4 categories in [insert year] – paediatric cases

CD4 categories	Total	
	N	%
Advanced		
Mild		
Missing		
Not Significant		
Severe		

Table H: Deaths reported among [number of patients] patients in [insert year], according to age and sex

Age group	Sex		Total	
	Male	Female	n	%
0-9				
10-14				
15-19				
20-24				
25-29				
30-34				
35-39				
40-44				
45-49				
50-54				
55-59				
60-64				
65+				
Total				
Median age at death (years)				

Table I: Cumulative reported HIV cases and HIV deaths from [insert date] to [insert date]

Sex	New HIV cases	HIV deaths
Male		
Female		
Total		

Table J: Annual number and rates of new reported HIV cases and HIV-related deaths from [insert year] to [insert year]

Year	HIV cases				Death among people with HIV			
	Male	Female	Total	HIV cases per 100,000	Male	Female	Total	Annual mortality rates among PLHIV
2014								
2015								
2016								
etc.								
Total								

Figure A: New HIV diagnosis in [insert year] by age and sex

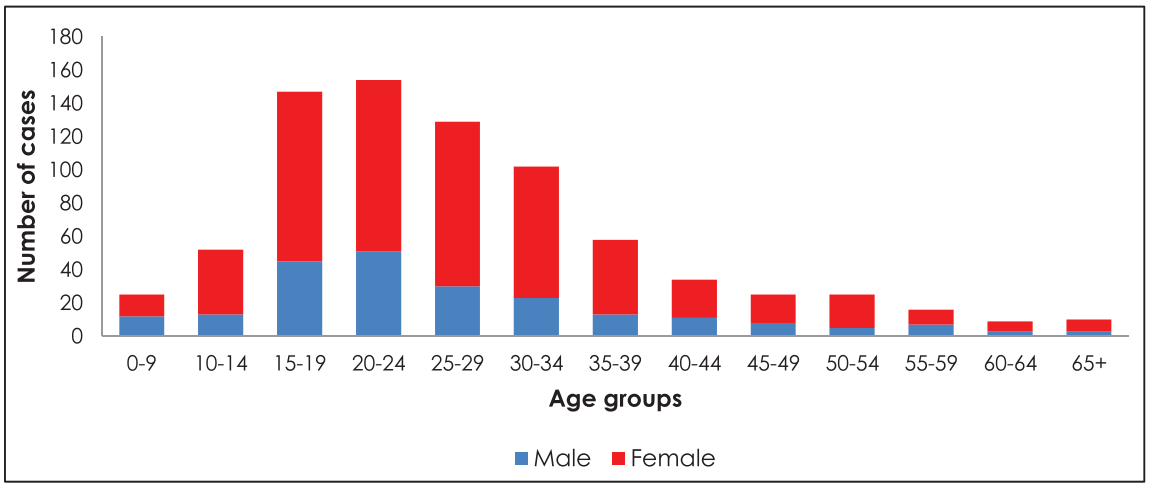


Figure B: Annual Trends of HIV cases and deaths

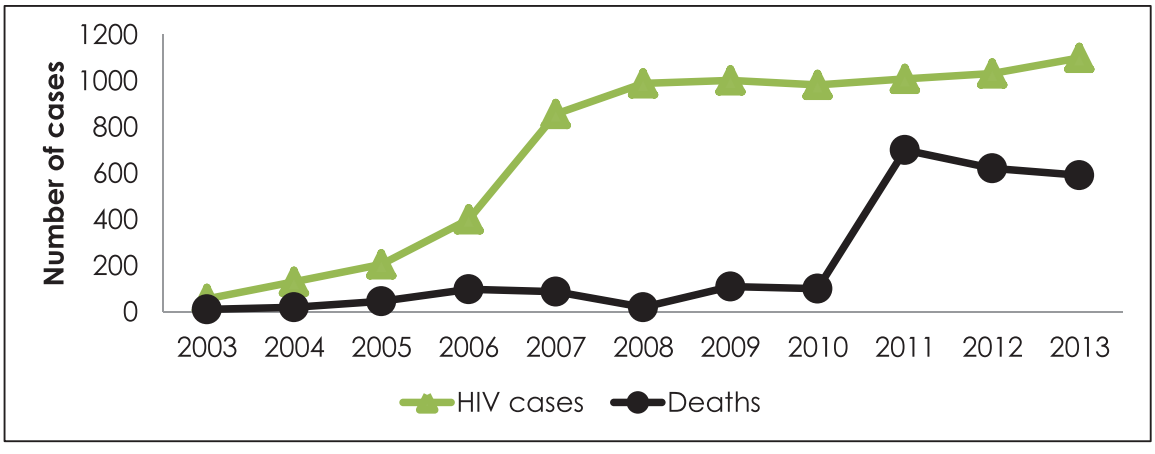


Figure C: HIV deaths by age and sex in [insert year]

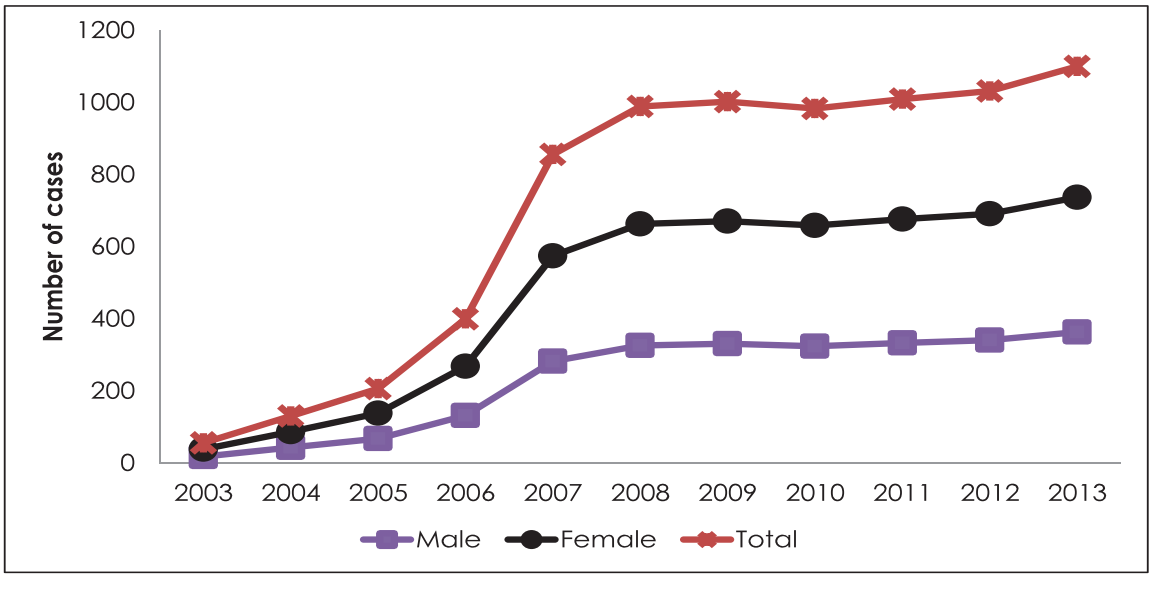


Figure D: Proportion of PLHIV alive at 12, 24 and 36 months after diagnosis, [insert year] – [insert year]

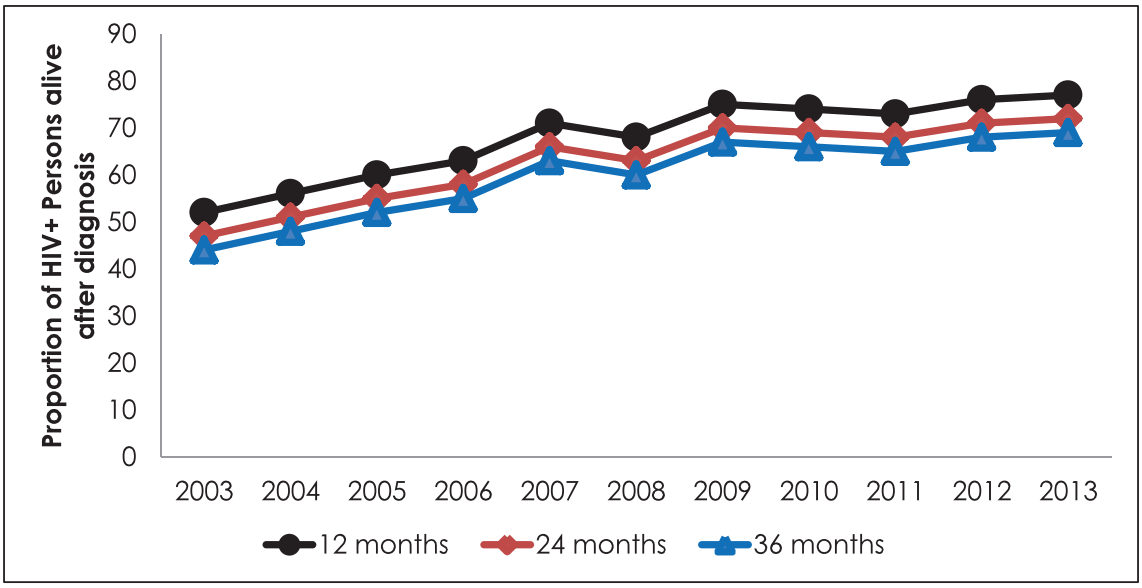


Figure E: Annual trend of new HIV cases by sex, [insert year] – [insert year]

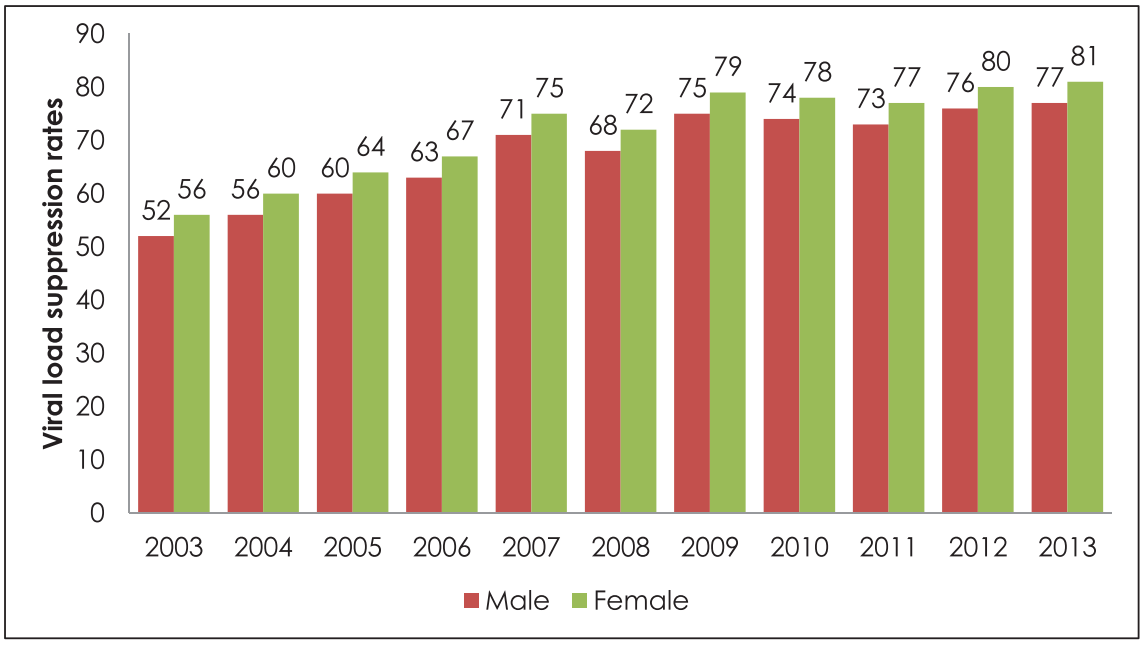
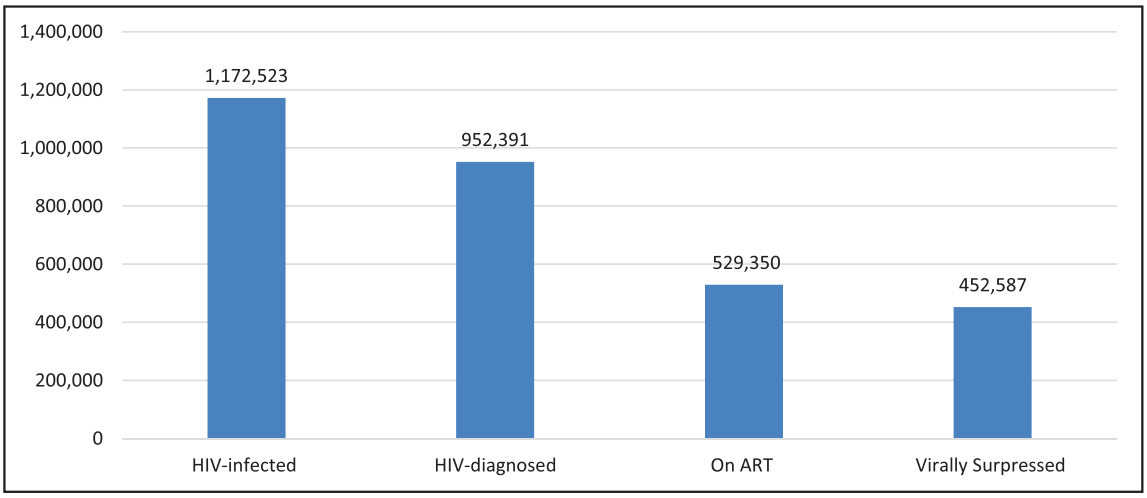


Figure F: HIV treatment cascade, [insert geographic location and year]



Annex VI: Sample CBS Dashboards

Figure A: Example of dashboard: Epidemic Profile

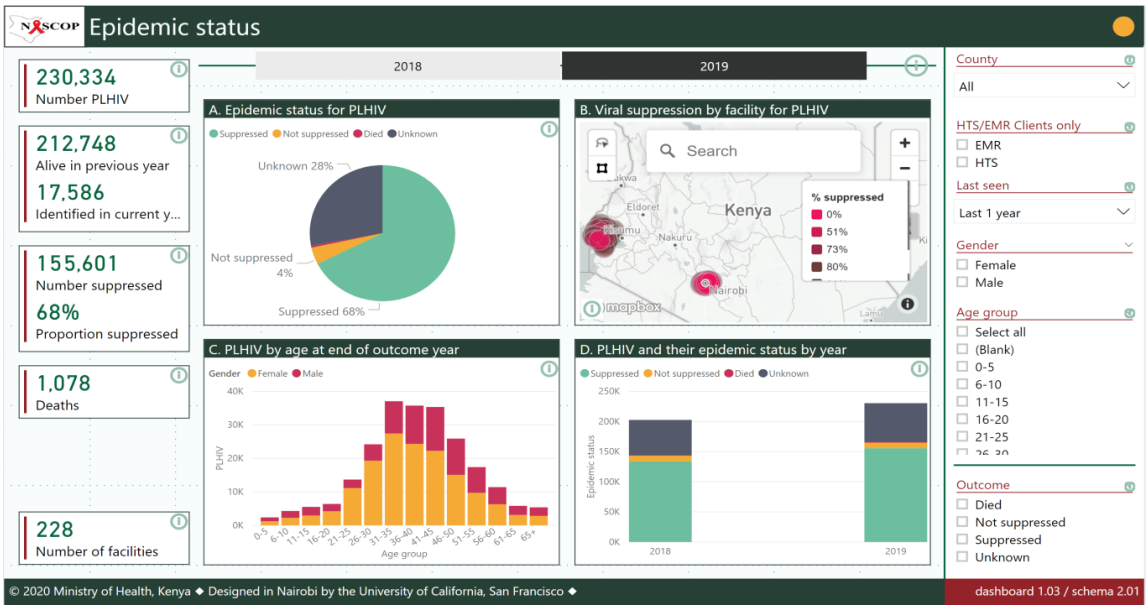


Figure B: Example of dashboard: New Diagnosis

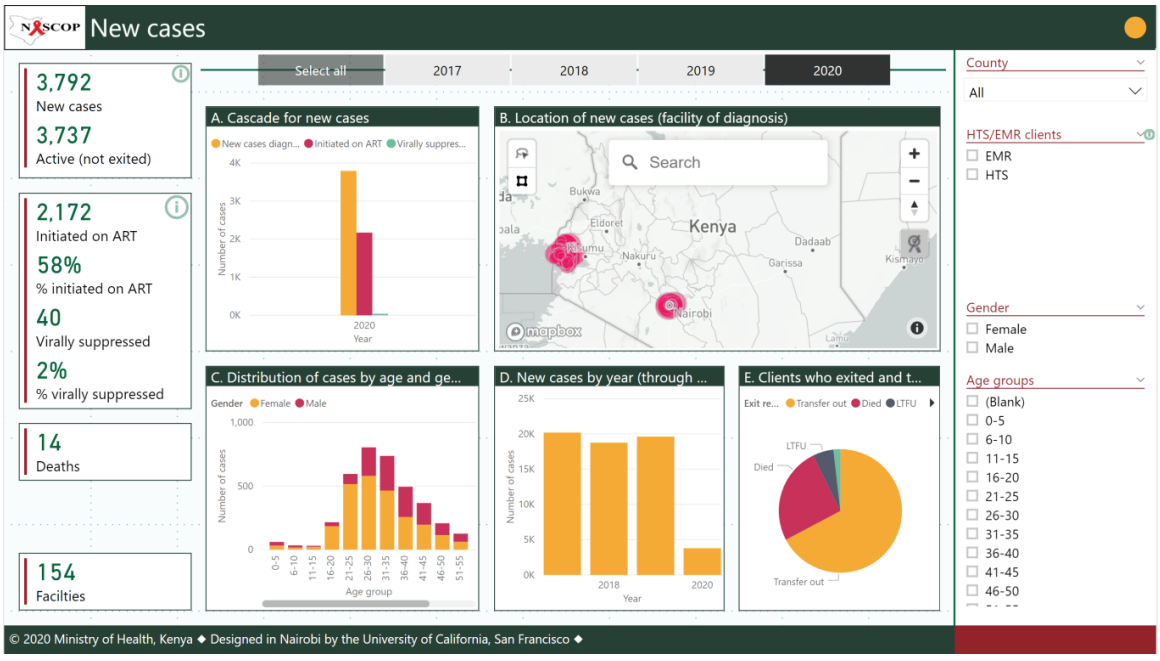
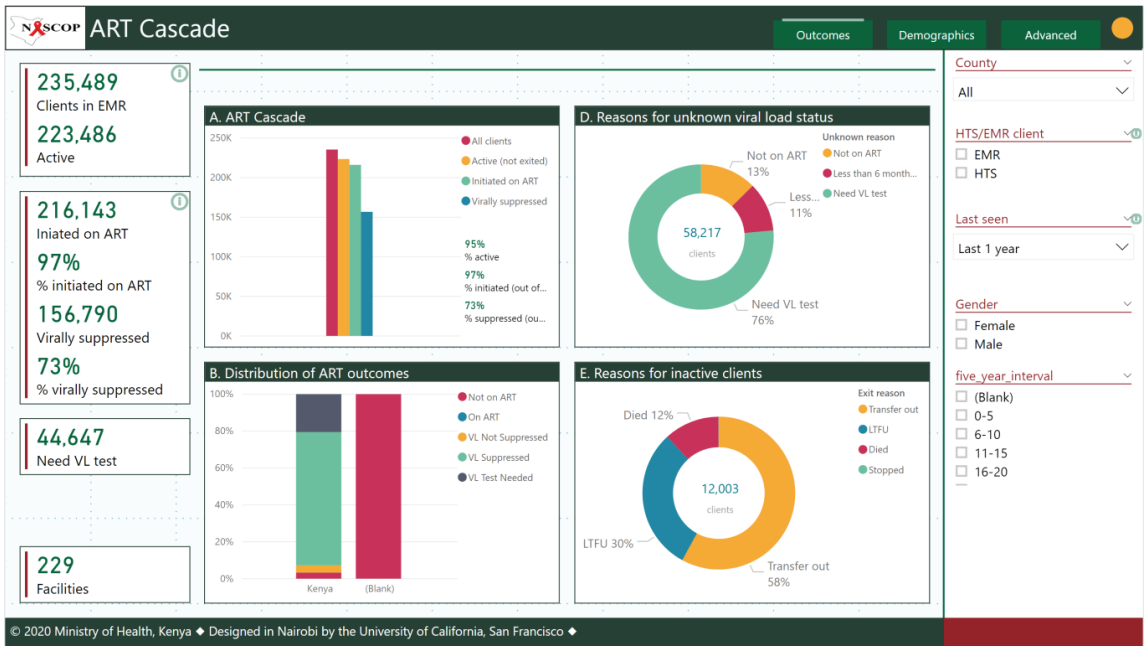


Figure C: Example of dashboard: ART Cascade



National AIDS and STI Control Programme (NASCOP)
Ministry of Health, Kenya
P.O Box: 19361, Code: 00202 Nairobi, Kenya
T: +254 (0) 20 2630867
E: headnascop.moh@gmail.com

NASCOP

KENYANS AND AMERICANS IN PARTNERSHIP TO FIGHT HIV/AIDS
PEPFAR

UCSF Institute for Global Health Sciences
Global Strategic Information