A TOOL FOR PREDICTING LOSS-TO-FOLLOW-UP AMONG PEOPLE LIVING WITH HIV AT BUSIA BORDER

BY

DENICE O. JUMA

MASTER OF SCIENCE IN DATA ANALYTICS

KCA UNIVERSITY

2017

A TOOL FOR PREDICTING LOSS-TO-FOLLOW-UP AMONG PEOPLE LIVING WITH HIV AT BUSIA BORDER

BY

DENICE O. JUMA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN DATA ANALYTICS IN THE FACULTY OF COMMUNICATION AND INFORMATION MANAGEMENT AT KCA UNIVERSITY

November, 2017

DECLARATION

I declare that this dissertation is my original work and has not been previously published or submitted elsewhere for award of a degree. I also declare that this contains no material written or published by other people except where due reference is made and author duly acknowledged.

Student name: **Denice O. Juma** Registration number: **15/05987**

Sign: _____ Date: _____

I do hereby confirm that I have examined the master's dissertation of

Denice O. Juma

And have approved it for examination

Sign: _____ Date: _____

Dr. Mwangi Henry

Dissertation Supervisor

Sign: _____ Date: _____

Mrs. Rachael Kibuku

Dissertation Supervisor

ABSTRACT

Human Immuno-Deficiency Virus (HIV) and Acquired Immune Deficiency Syndrome (AIDS) are a global emergency. Infection with HIV can lead to poor health, loss of lives, increased number of orphans and reduced national productivity. In Kenya and Uganda, National AIDS & STI Control Programme (NASCOP) and the Uganda AIDS Commission (UAC) respectively are promoting combination of approaches for HIV prevention with the key populations. Studies have revealed that 1.5 million people live with HIV with a 5.9% adult HIV prevalence. There are an estimated 78,000 new HIV infections with 36,000 aids related deaths and 59% adults on antiretroviral treatment (AVERT 2016). Complex socio-cultural, economic, and health-system factors inhibit excellent patient retention. Better tracking, enhanced social support, and regular adherence counselling in addressing stigma, and alternative healing options are needed. Intervention strategies aimed at changing clinic routines and improving patient–provider communication could address many of the identified barriers (Tiruneh *et al*. 2016). The objective of the study is to develop a tool to predict possible loss-to-follow-up among mobile people living with HIV/AIDS enrolled in care and treatment at the Busia border.

Adherence to ARV drugs and retention in care and treatment programs of a sample of PLHIVs in the Northern Transport Corridor of Kenya and Uganda particularly in Busia cross-border site was assessed to determine the factors leading to loss-to-follow-up among them. A design science methodology was adopted in the design, development, testing, implementation and validation of the data mining and analytics tool for predicting possible loss-to-follow-up among people living with HIV. The tool was piloted at two cross-border participating facilities close to the border crossing point. Missing variables speaking to nationality and cross-border and cross-county/district mobility characteristics were collected through a community and facility profile form. Exiting data from daily activity registers were filtered based on main dependent LTFU predictor variables resulting from reviewed literature on factors, used to inform design of the tool and as the training dataset. New PLHIV were enrolled and their data run through the decision tree predictor for results.

Metrics were used to assess the differences in tracking and management of PLHIVs before and after implementation of the tool. Model evaluation metrics was used to test the accuracy, efficacy and utility of the tool. Study findings inform future effective interventions on tracing and linking back mobile PLHIVs displaying transnational service access characteristics to adhere and be retained in care and treatment as an effort to achieving the global UNAIDS 90-90-90 targets.

Keywords: data mining, loss-to-follow-up, mobility, cross-border referral

ACKNOWLEDGEMENT

I acknowledge the support of my supervisors Dr. Mwangi Henry and Mrs. Rachael Kibuku, colleagues, and other lecturers for their unwavering support and immense contribution and guidance. My family and closest friends whose help and emotional support has seen me through my lowest moments. And above all God, the creator, giver of all that is good for seeing me through this process.

TABLE OF CONTENTS

DEDICATION

This dissertation work is dedicated to my future wife Patience Wakesho, who has been a constant source of support and encouragement during the challenges of school and life. I am truly thankful for having you in my life. This work is also dedicated to my parents, Pamella Juma and John Juma, who have always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

AIDS:	Acquired Immune Deficiency Syndrome
ART:	Anti-Retroviral Therapy
ARVs:	Anti-Retroviral Viral
CCC:	Comprehensive Care Clinic
HRIO:	Health Records and Information Officer
CHV:	Community Health Volunteers
DHIS:	District health information system
KII:	Key Informant Interview
KP:	Key Population
LDTDs:	Long Distance Truck Drivers
MOH:	Ministry of Health
NACC	National AIDS Control Council
NASCOP	National AIDS and STI Control Programme
SRH	Sexual Reproductive Health
TB	Tuberculosis
VCT	Voluntary Counselling and Testing

OPERATIONAL DEFINITION OF TERMS

1. **Risk:** the probability that a person may acquire HIV infection, usually because of specific behaviours that enable HIV transmission to occur (UNAIDS 2007)
2. **Discrimination:** Discrimination is treating someone in a different prejudicial manner, often by their belonging, or being perceived to belong to a group. It is the unfair, unfavourable and unjust treatment based on prejudice regarding race, colour, sex, perceived HIV status or diseases characterised by rejection, isolation, and discrediting an individual.
3. **HIV-related Stigma:** Real or perceived negative feeling to a person or group of individuals (in this case PLHIV) by HIV-positive status. Someone is ignored, socially excluded and treated differently from others because of their HIV status.
4. **Immunodeficiency**: (or immune deficiency) is a state in which the immune system's ability to fight infectious disease and cancer is compromised or entirely absent. Most cases of immunodeficiency are acquired ("secondary") due to extrinsic factors that affect the patient's immune system.
5. **Truckers**: For the purpose of this study, truckers (including turn boys) are those who drive trucks more than 400 kilometers in either direction or are away from the originating points for more than one night and having significant risk for HIV.
6. **Transport Corridor**: This is the transport corridor linking the Great Lakes countries of Burundi, Democratic Republic of Congo, Rwanda, and Uganda to the Kenyan seaport of Mombasa. In this context, it is linking Mombasa to Busia district/county - border site of Kenya and Uganda.

7. **Sexual violence**: This is the act of physically forcing or psychologically intimidating a person to engage in sex or subjected to sex acts against their will that one finds degrading or humiliating.
8. **Risk**: the probability that a person may acquire HIV infection, usually because of specific behaviors that enable HIV transmission to occur (UNAIDS 2007).
9. **Vulnerability**: a state when an individual's ability to avoid infection is diminished by inadequate personal knowledge or skills, by cultural norms that validate risky behaviors, or by circumstances that make risk reduction difficult or impossible (UNAIDS 2007).
10. **De-identified information:** All patient information entered into a database using coded identification numbers, and no information that could reveal patient identity was entered into the database.
11. **Lost to follow - up**: is not keeping ART refill appointment for 90 days or longer from the last booked refill appointment date, yet not classified in the patient clinical outcome as 'dead' or 'transferred - out'.
12. **Confirmed lost to follow - up**: refers to a HIV positive individual who is not keeping ART refill appointment for 90 days or longer from the last booked refill appointment date, and being found alive but not on ART.
13. **True lost to follow - up**: means not keeping ART refill appointment for 90 days or longer from the last booked refill appointment date, and being found alive but not on ART. Or not being traceable to allow confirmation of final clinical outcomes such as 'dead' or 'transferred - out.'

CHAPTER ONE
INTRODUCTION

1.1 Background of the Study

Human Immunodeficiency Virus attacks a vital part of the immune system, the CD4 cells that fight infections and disease, making the human body unable to fight infections. The body gets Acquired Immunodeficiency Syndrome (AIDS) and becomes prone to opportunistic infections (OIs) that may cause death (Murphy *et al*. 2008). An estimated 36.7 million (33.2 million–37.2 million) people globally were living with HIV as of 2016. The number of individuals newly infected with HIV, especially children, and the number of AIDS-related deaths has declined over the years and the number of people with HIV receiving treatment increased to more than 18 million in 2016 (Kaiser Family Foundation 2016). Ninety-seven percent of the proportion living with HIV reside in low and middle-income nations (Kerrigan *et al*. 2003). Ten countries in Sub-Saharan Africa, three of which are East African Community (EAC) Partner States - Kenya, Uganda and the United Republic of Tanzania, account for 81% of all people living with HIV in the region (ONUSIDA 2013).

The prevalence rate for HIV in Kenya among adults aged 15-64 years is 6.0% (National AIDS Control Council 2014) whereas that of Uganda is 6.4% (AIDS Commission 2009). The Kenya HIV Obviation Replication and Modes of Transmission Analysis (2009) designates that 93%, 70% of HIV transmission in Kenya, Uganda respectively occurs sexually, both in the context of transactional and commercial sex and in longer-term relationships, including marriage (Gelmon *et al*. 2009). Also, the Kenya national guidelines for HIV/STI programs for sex workers (2014), shows that the country is experiencing a generalised epidemic; however, different dynamics and drivers exist among certain populations increasing their HIV risk. Estimates by the Joint United

Nations Programme on HIV/AIDS (UNAIDS) suggest that as many as 50% of all new HIV infections worldwide occur in people from key populations (Kaiser Family Foundation 2016). The UNGASS reports indicate that the risk behaviours and vulnerabilities of the key populations result in there being tremendously affected by HIV compared to the general population.

According to Kenya AIDS Strategic Framework (KASF), 2014/2015 – 2018/2019 and Uganda National HIV And AIDS Strategic Plan 2015/2016- 2019/2020, these populations are at significant risk of transmitting and acquiring HIV/STIs. It is mainly due to increased frequency of high-risk sex and that they have great potential to transmit the virus through their extensive network of sexual partners (NACC 2014), (An AIDS-free Uganda & Uganda AIDS Commission 2015). A significant percentage of female sex workers in Kenya, 29.3%, are living with HIV (NASCOP 2014) and together with their clients, they contribute 34% percent of all new HIV infections (National AIDS Control Council 2014), hence the need to intervene through FSWs to reduce the incidence of HIV in the country. Kenya's HIV Prevalence is at 6.1%, whereas that of gender is at 3.5%, 1.5% for female and males respectively (NASCOP 2014). It is disproportionate in the country in the Western region, Migori, Homabay County – Muhuru Bay –termed as hyperendemic with a high 15% prevalence. The prevalence of Key Populations (KP) is five times greater than the overall countries with Female Sex Workers at 29.3% followed by People Who Use Drugs at 18.3% and lastly, Men having Sex with Men at 18.2%. Comparatively, as of 2015, the estimated HIV prevalence among adults (aged 15 to 49) in Uganda stood at 7.1% (UNAIDS 2016). The future is looking brighter since as of 2015 around 40% of adults living with HIV were still not on treatment. Persistent disparities remain around who are accessing treatment and many people living with HIV experience stigma and discrimination (An AIDS free Uganda & Uganda AIDS Commission 2015).

TABLE 1
Demographics comparison for Kenya and Uganda (Index Mundi Website)

Country	**Kenya**	**Uganda**
Population	45,010,056	35,918,915
Growth rate	2.11% (2014 estimate)	3.24% (2014 estimate)
Birth rate	28.27 births/1,000 population (2014 estimate)	44.17 births/1,000 population (2014 estimate)
Death rate	7 deaths/1,000 population (2014 estimate)	10.97 deaths/1,000 population (2014 estimate)
Migration rate	-0.22/1,000 population (2014 estimate)	-0.76/1,000 population (2014 estimate)
Fertility rate	3.54 children born/woman (2014 estimate)	5.97 children born/woman (2014 estimate)
Prevalence	6.1% (2012 estimate) - Adults	7.2% (2012 estimate) - Adults
PLHIVs	1.646 million (2012 estimate)	1,549,200 (2012 estimate)
HIV deaths	57,500 (2012 estimate)	63,300 (2012 estimate)

Source: (CIA Fact book 2015, Index Mundi Demographics Comparison)

Along the transport corridor, FSWs are mostly mobile and engage in sex primarily through bars or lodging, and they experience poverty and lack of opportunity compounded by factors including high frequency of inconsistency of condom use creating a risk environment making them vulnerable to the transmission of HIV (Fuller *et al*. 2009). The National Strategy on HIV and AIDS and STI programming along transport corridors, Kenya indicates that sexual violence, low levels of education, stigma, discrimination, and age of entrance into the profession, low health seeking behaviour, as well as general lack of health access also contribute to the risk environment. (NACC and NASCOP 2012).

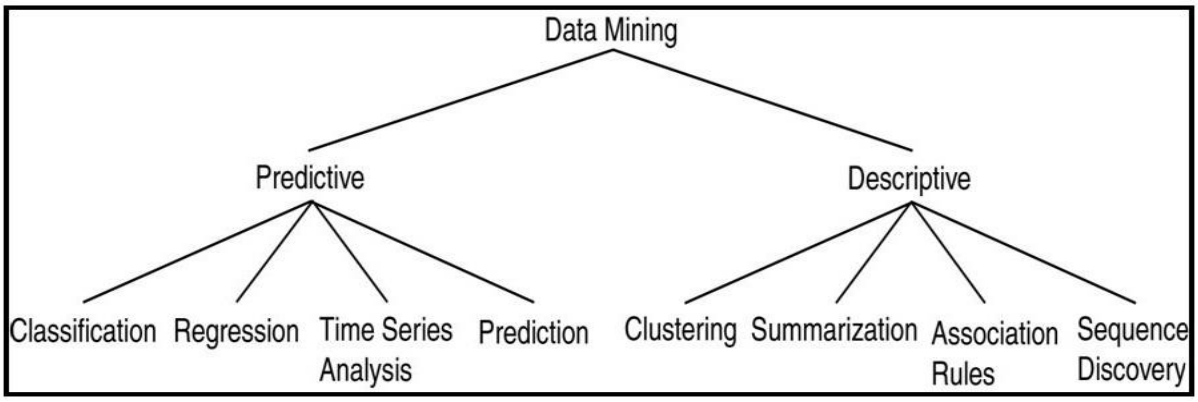
There is a growing concern about the increasing rates of Loss to Follow-up (LTFU) among people who are on HIV/AIDS treatment programs. Defaulting ARV drugs is more common in resource-poor settings, little is known about the time to LTFU and predictors after Antiretroviral Therapy (ART) initiation in cross-border settings (Tadesse & Haile 2014). Long-term regular follow-up of ART is an important component of HIV care. Patients who are lost to follow-up

(LTFU) while on treatment compromise their health and the long-term success of ART programs (Berheto *et al*. 2014). A better understanding of the risk factors associated with attrition could be helpful to design interventions to reduce mortality and LTFU in patients who initiate ART (Alvarez-Uria *et al*. 2013). The number of PLHIVs meeting the threshold of being enrolled in ARV drugs is going up. Specialty programmatic accomplishment calls for that PLHIVs who're taking ARV medications stay on medicine, what's more, are watched over consistently (Mberi *et al*. 2015). In settings with limited resources, first adherence patterns, as well as CD4+ cell counts and unsuppressed viral load, at any time point in treatment are predictive of loss and serve as efficient markers for developing targeted interventions to reduce rates of attrition (Meloni *et al*. 2014). To improve retention in ART programs for HIV control, counselling and testing among priority populations, should focus intensified intervention targeting those who were more likely to be LTFU, the young, less educated, unofficial residents with smaller social network size (Tang *et al*. 2015).

Data mining is a process which finds useful patterns from a large amount of data. Data mining techniques provide people with the new power to research and to manipulate the current large volume of data. One of the aims of data mining can be seen as the analysis of observational datasets to find unsuspected relationships and to summarise the data in novel ways that are both understandable and useful. This relationships and summaries derived through data mining are often referred to as models or patterns (Hailu 2011). The Data Mining Process involves selecting the target data, preprocessing the data, transforming them if necessary, performing data mining to extract patterns and relationships, and then interpreting and assessing the discovered structures (Hand *et al*. 2001). There are two main Data Mining Models. The predictive model makes a

prediction about unknown data values by using the known values. The descriptive model identifies the patterns or relationships in data and explores the properties of the data examined.

FIGURE 1
Data mining models and tasks



The massive amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analysed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision-making. Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarise the data in novel ways that are both understandable and useful to the data owner (Dey & Rautaray 2014). One of the underlying concerns of health care providers is the reducing defaulter rates at cross-border areas, besides existing functional national systems.

Data mining it is one way to display an algorithm; specifically, decision trees are a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. Classification algorithms are widely used for customer retention or to develop a recommender system. Decision trees are efficient when there is a linear decision boundary like classifying people on the basis of their IQ. It fits in my study as a technique for predicting possible LTFUs from existing daily activity register data and

adherence to drug refill/treatment appointments by PLHIVs at the selected cross-border sites. Independent and dependent variables that must be present backed by literature review and lessons learnt over time were used in developing the predictive tool.

The message from the latest joint cross-border KP Technical Work Group (TWG) was that tracking of KPs is below optimum and that partners need to put more effort in referrals and linkages of PLHIVs for care, treatment, and viral load suppression. Unique identifier codes are not yet institutionalised. Currently, there are no collaboration and communication between health facilities across the border; there is a need for a harmonised referral forms and HCWs trained and sensitised on its use, organisations running HIV programs are using disparate in-country silo systems for management and tracking of LTFUs. The HMIS do not collect nationality and mobility information of the clients receiving services at the health facilities. Therefore, a client's profile form with appropriate job aids should be made available at service provision units to capture this vital information.

1.2 Statement of the problem

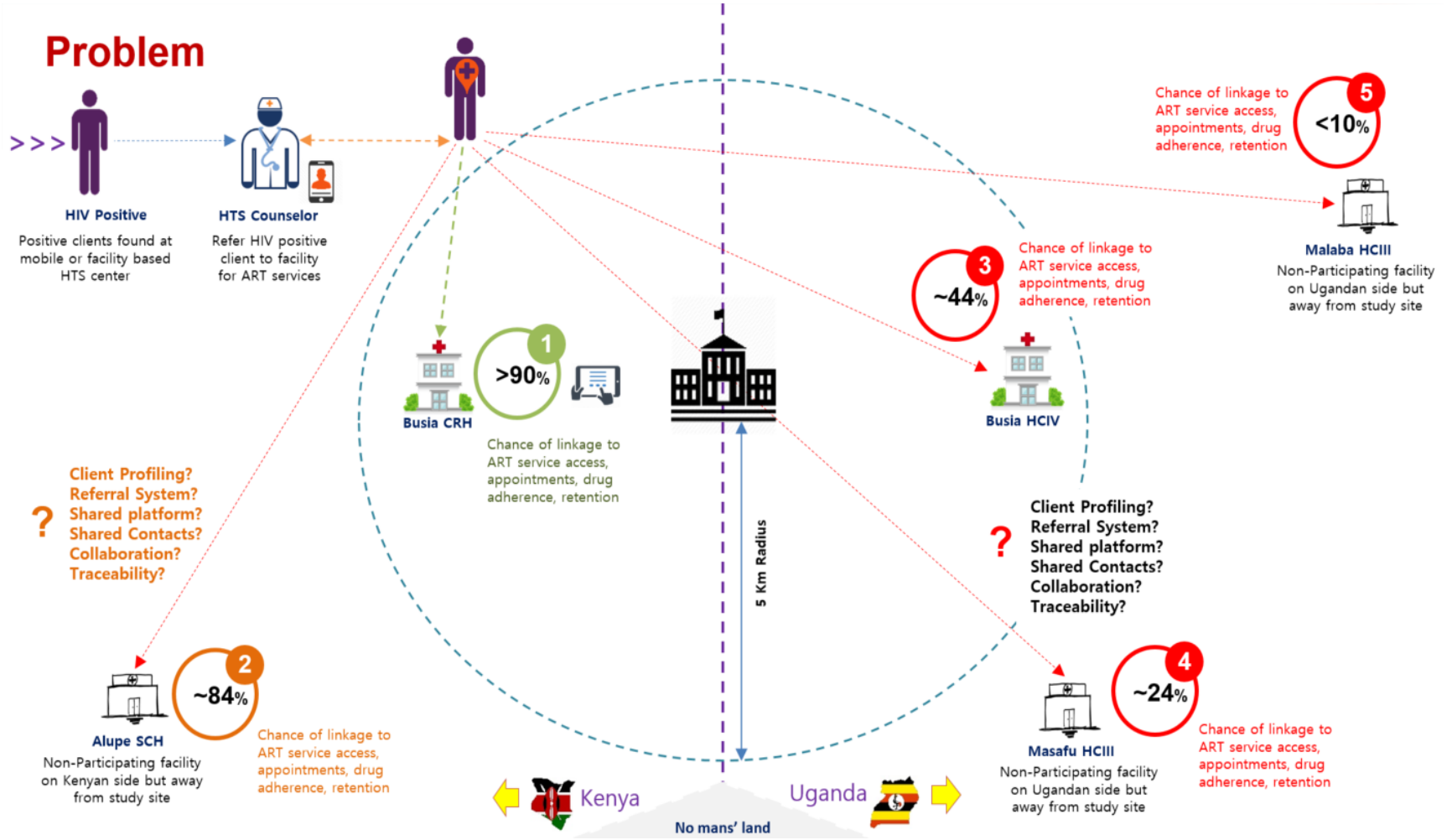
The high early death rates indicate that HIV-infected individuals from mobile populations within the cross-border sites are enrolling into Anti-retroviral Therapy with far too advanced immunodeficiency. They visit health facilities across the border citing reasons like; stigma, discrimination, the disclosure of their status, distance from the facility, attitude of the HCPs, service availability, but do not get services and the support they need due to higher service costs subjected to non-nationals, lack of completed referral forms acceptable across the border for treatment, lack of coordination and communication between health care providers working at health facilities across borders. To the researcher's knowledge, there is no data to demonstrate how Kenyans receive services on the Ugandan side and vice versa. It is also a well-known fact that

globalisation has increased both legal and illegal movements tremendously across borders. As a result, fueling the need for strategies to ensure the continuity of treatment and care of PLHIVs moving across borders (Meloni *et al*. 2014).

Treatment for mobile populations is complicated since many are unable to remain in each location long enough to complete the lengthy ART regimen. The main challenges include PLHIVs defaulting drugs due to failure to report to the health facility they are enrolled in ART when they move from the Kenyan to Ugandan hotspots and back, stigma and discrimination, lack of proper communication and feedback mechanism between cross-border referral facilities, a lacking transnational referral system and nonexistence cross-border communication channel. The significant gaps include; information sharing and coordination mechanism between countries on cross-border referral, efficient referral mechanism between in-country health care settings and the population at the sites and lack of cross-border coordination mechanisms (Tiruneh *et al*. 2016).

Despite the efforts by organizations to have their data collection subsystems (DHIS2, KEMSA LMIS, AIDSInfo, AMPATH POC, SANITAS) for managing data used for program planning and evaluation, there are no cross-border referral systems for tracking and monitoring mobile PLHIVs who enroll for care in another country and take ARV drugs across the border and therefore referred to as a defaulter in the country/facility they enrolled in care. These necessitate inter-country policy discussions for the development of cross-border referral tools for predicting and tracking mobile PLHIVs on ART. The linkage between transnational peer educators who mobilise their peers to access HTC services, the health care providers (HCPs) at the sanitation facilities who diagnose the patients and confirm their HIV-positive status (The National ART Program Swaziland 2012).

FIGURE 2
Proximity to referral facility and across-borders



Additionally, the community health workers (CHWs/CHEWs) who follow-up the PLHIVs enrolled in care to ensure they adhere to ARV drugs and the VHTs/linkage facilitators working through PLHIV support groups is an end to having an efficient system. Joint cross-border meetings by the C/DHMTs on health issues and exchange visits, communication and collaboration between the HCPs through a shared contacts directory, availability of well-recognized client profiling and inter-facility referral forms together with the willingness of the HCPs to complete them. The efficacy of the referral system is the potential to strengthen the limited in-country tracking of mobile PLHIVs and the ability to ensure adherence to ARV drugs thus reduce deaths due to HIV/AIDS (Tiruneh *et al*. 2016).

1.3 Purpose of the Study

The goal of this study is to understand the factors that motivate mobile PLHIVs to default anti-retroviral drugs (ARV), understand the causes or effects of these factors to overcome the difficulty with increasing rates of loss-to-follow-up at two participating health facilities at the cross-border study site. The study seeks to discover what is required in designing a tool that utilizes data mining techniques to mine existing PLHIV daily activity register data to predict if a client would be a possible loss-to-follow-up. This has the potential of refining our current understanding of available HMIS that to a large extent serve as stores for PLHIV data.

Consequently, the researcher also seeks to determine what is required in developing a data mining and analytics tool to predict possible loss-to-follow-up among mobile PLHIVs which were then validated and implemented at the facilities and thus provide a new interpretation of tracing mobile PLHIVs displaying transnational service access characteristics. This would contribute to the understanding of what makes adherence to ARV drugs and retention of enrolled PLHIVs to care/treatment successful or unsuccessful.

1.4 Research Objectives

1.4.1 Main objective

The general objective of the study is to develop a tool for predicting loss-to-follow-up among people living with HIV at the Busia border

1.4.2 Specific objectives

i. To identify the factors that motivate people living with HIV to default anti-retroviral drugs (ARV) and the levels of loss-to-follow-up
ii. To design a data mining and analytics tool to predict possible loss-to-follow-up among people living with HIV
iii. To develop a data mining and analytics tool to predict possible loss-to-follow-up among people living with HIV
iv. To validate and implement the data mining and analytics tool for predicting possible loss-to-follow-up among people living with HIV

1.5 Research questions

i. What are the factors that motivate people living with HIV to default anti-retroviral drugs (ARV) and what are the levels of loss-to-follow-up?
ii. What should be considered in designing a data mining and analytics tool for predicting loss-to-follow-up among people living with HIV?
iii. What should be considered in developing a data mining and analytics tool for predicting loss-to-follow-up among people living with HIV?
iv. Which approach should be used to validate and implement the data mining and analytics tool for predicting possible loss-to-follow-up among people living with HIV?

1.6 Study significance

The study enables HIV/AIDS policy makers to come up with policies that support the tracking of PLHIVs who are loss-to-follow-up. The prevalence and incidence estimates by the National AIDS Control Council (NACC) are more accurate as the mobile populations at the cross-border sites which have in most instances been left out. Through macro planning, the county/national governments plan and consider Kenyans or Ugandans who are receiving services at cross-border health care facilities.

The findings of this study may help them to design a special intervention program or improve the existing one. As a result, the population might be benefited if the service is improved or a new program is designed and implemented based on the output of this study. Currently, there are no cross-border referral systems in existence for tracking mobile LTFUs. The findings of this study are to be scaled up to potential cross-border sites within the East African Community (EAC) region and its recommendations used to inform future studies. The implementation and operationalization of the cross-border referral system strengthens the work of the HCPs.

It encourages the collaboration and coordination of the cross-border health facilities by providing a platform where they can share experiences and findings. The adoption of the inter-facility referral tool is an effort in harmonising the referrals processes. For researchers, the study can contribute to how the application of data mining is helpful in predicting possible LTFUs and identifies determinants of LTFUs through different algorithms. Hence, it can also invite interested researchers to explore more in related and similar areas for the future. Similarly, the benefits of LTFUs prediction can be seen at the individual, community and population levels.

1.7 Motivation

It is hoped that the results of this research such as the proportions of the non-nationals that seek services at cross-border health facilities, the ability to refer clients across borders through the referral system together with the lessons learned in implementing a model for tracking LTFUs informs and guide cross-border program direction and implementation. Lost- to- follow- up patients located are encouraged to return to the health facility for step- up adherence counselling and possible re- initiation on ART if willing to re- enroll into care. The provision of ART to migrant populations raises particular challenges on ensuring adequate treatment support, adherence, and retention in care. Thus a better understanding of individuals more likely to drop-out and has the potential to empower outreach for this key population. Providing ART to migrant workers is feasible, and highlights the need for the development and adaptation of models of care for migrant populations to ensure continuity of ART. Such support should extend beyond the early phase of ART initiation which in is considered as the risk period for defaulting from care.

1.8 Justification of the study

Steps in Profiling mobile clients through the introduction of nationality, target population and mobility variables in the national tools, harmonized inter-country referral tool, a shared platform for managing ART data, and a contacts directory of CBHUs has the potential to efficiently track LTFUs seeking services across the border and re-initiate them back for step- up ART adherence. This can generate meaningful data to inform policy decisions as statistics on defaulter rates reported by MOH are more accurate. Providers currently do not report mobile PLHIVs moving to neighboring countries for ART services due to stigma and confidentiality issues as true LTFUs when they are still alive but not on drugs. The study enables HIV/AIDS policy makers to come up with policies that support the tracking of PLHIVs who default drugs and end up reported as loss-

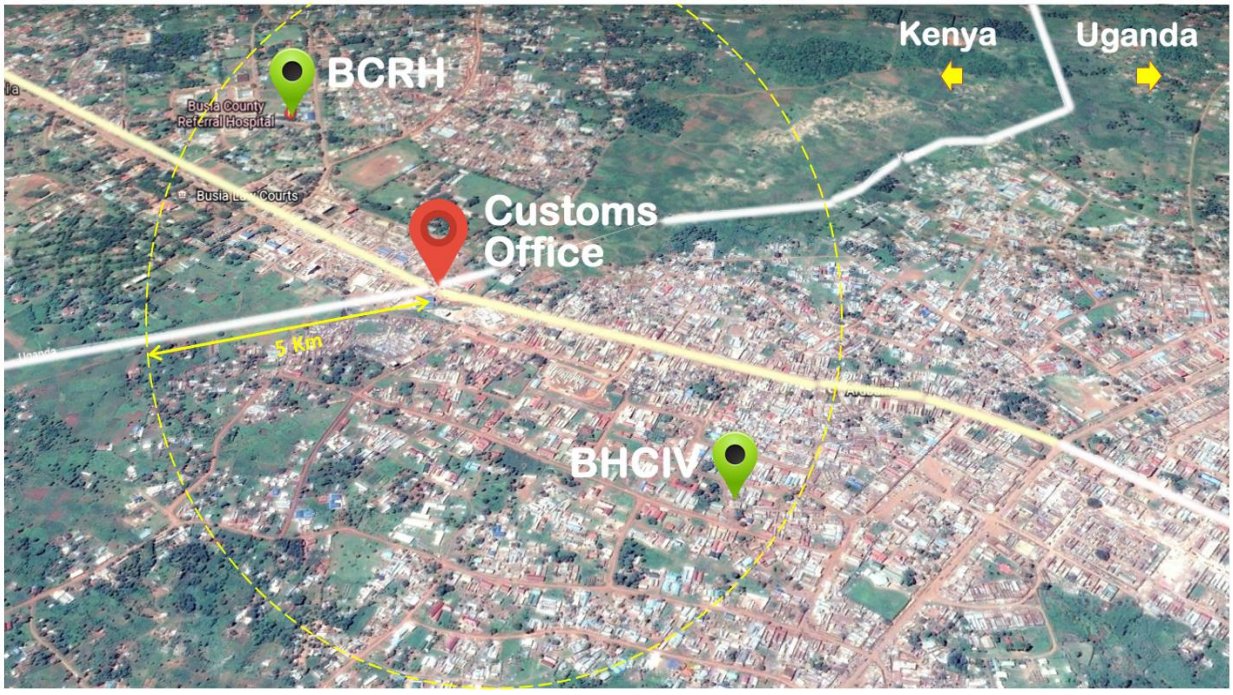
to-follow-up within districts/counties and across national borders. This tool assists the Kenyan and Ugandan County/District governments through their respective health ministries to reach the global UNAIDS 90-90-90 Target of testing to know HIV status, enrollment in care and retention/adherence. The prevalence and incidence estimates by the NACC (Kenya) and UAC (Uganda) are more accurate as mobile and migrant populations at the cross-border catchment areas have in most instances been left out. Through macro planning, the county/national governments strategize and consider Kenyans or Ugandans who are receiving services at cross-border health care facilities (transnational service access). The findings of this study may help programs to design a special intervention or improve the existing ones. As a result, the population might benefit if the service is improved or a new program is designed and implemented based on the output of this study. Currently, there are no cross-border referral systems in existence for tracking mobile LTFUs. The findings of this study can be scaled up to approved cross-border sites within EAC region and its recommendations used to inform future studies

1.9 Scope of the Study

- This study explores the growing concern about the increasing rate of loss-to-follow-up (LTFU) amongst mobile people living with HIV/AIDS (PLHIVs) at Busia border that leads to HIV-related deaths.
- The study site is an area within a radius of 5 kilometres from the border crossing point of Busia County (Kenya) and Busia District (Uganda) along the main transport corridor. The Criteria for the selection of the study site include: (1) Cross-border movement and trade, (2) High HIV/STI prevalence, (3) Gaps in services, (4) Presence of key and vulnerable populations including migrants (5) Recognized by EAC and partner states as a priority underserved cross-border area and (6) Underserved by bilateral programs.

- The target population are mobile (frequently cross the border or County/District in a country) PLHIVs enrolled in care and treatment at two public health facilities; Busia County Referral Hospital (Kenya) and Busia Health Centre IV (Uganda).
- True loss to follow up is defined in this study as "an infected individual who is not keeping ART refill appointments for 90 days or longer from the last booked refill appointment date, and being found alive but not on ART drugs, or not being traceable to allow confirmation of final clinical outcomes such as 'dead' or 'transferred - out".
- The researcher intends to study the reasons for these retention/adherence gaps by using a design science methodology where the researcher utilized mixed methods to implement a model for predicting and monitoring possible loss-to-follow-up.

FIGURE 3
Land cross-border learning site



Source: CB-HIPP Site Mapping

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

For this study, the literature was examined to gain more understanding of the main factors that motivate mobile, vulnerable PLHIVs to default ARV drugs, the levels of LTFUs among the mobile population at the participating health facilities. The main challenges faced by HCPs at the in following up LTFU cases, the establishment of an appropriate model for predicting LTFUs among vulnerable mobile priority populations at cross-border sites and associated factors. The review covers theoretical and empirical sources related to the main concepts found in this study. For this study, a literature review was conducted the before data collection and consulted documents acquired from various sources.

2.2 Theoretical Review

The different types of academic sources used for the study were categorised into research methodology sources and conceptual and theoretical sources. Sources on research methodology reviewed to refine the research design and processes for this study mostly included textbooks on research methodology and journal articles. Sources on conceptual and theoretical references included books on health education and health promotion theories and sociology. These resources were essential for identifying and refining concepts relevant to predicting and tracking ARV drug defaulters at cross-border sites.

The following assumptions are the starting points for this study:

- Lack of comprehensive service package and standard delivery mechanisms is a barrier to engage and retain HIV-positive patients in care
- Socio-economic and cultural norms can negatively affect retention of PLHIVs in care
- Patient's demographic characteristics (gender, age, marital status, education level) influence their perception of adherence to care services
- A significant percentage of patients enrolled in care become lost-to-follow-up before they are initiated on ART
- Most of these patients die; some of them return to care for severe illness and compromised immunity.
- Lost-to-follow-up patients continue to be the source of new HIV infection in the community as they are less likely to either disclose their HIV serostatus or adhere to preventive behaviours.

2.3 Empirical Review

2.3.1 Primary Sources

The literature review was based mainly on primary empirical sources contained in scientific journals and textbooks. In this review, it is clear that while there are many studies conducted on patient retention and factors associated with lost follow-up for patients on ART, there are few studies carried out on data mining models for predicting and tracking cross-border patients during the ART period. The few studies on predicting, proper referrals and linkages, retention of PLHIVs during the ART period and the suppression of their CD4 counts/viral load are from Malawi, South Africa, and Ethiopia.

2.3.2 Secondary sources

Secondary sources were also consulted not only to support primary sources but also to help in the identification of additional primary studies from ART care and treatment reviews and the use of data mining approaches in predicting possible loss-to-follow-ups at cross-border sites. These sources included program (briefs, bulletins, narratives) reports, technical and policy-related publications from the global (UNAIDS, WHO, MEASURE Evaluation, AVERT) and local (NASCOP, NACC, MOH, EAC, CB-HIPP and many others) institutions.

2.3.3 Hypothesis

H01: The hypothesis of this study is that the prediction model will not improve adherence of people living with HIV to ART.

After validation and implementation of the data mining and analytics tool for predicting possible loss-to-follow-up among people living with HIV, we expect to see more improvement in tracing PLHIVs, retention, adherence rates in the two participating facilities than in other non-participating facilities.

2.4 Related Work

2.4.1 Factors that motivate people living with HIV to default ARV drugs and the levels of loss-to-follow-up.

As antiretroviral treatment cohorts continue to expand, ensuring patient retention over time is an increasingly important concern. Together with capacity and human resource constraints, has led to the consideration of out-of-clinic models for the delivery of antiretroviral therapy (ART) (Decroo *et al*. 2011). According to the national guidelines of anti-retroviral therapy, the reasons for becoming LTFU can include stigma/discrimination, prefer traditional medicine, partner doesn't want patient to get ART, health improved, side effects, did not like the way they were treated at the clinic, inconvenient clinic hours, waiting time at the hospital, transport costs, being too sick to walk, didn't know to come back or other. High tracking costs, a lack of sufficient information on patient locations and a shortage of human resources necessary for finding missing patients have contributed to a backlog of patients needing to be traced**.**

Some studies have demonstrated a significant association of non-adherence with certain socio-demographic factors such as low literacy or education, lower individual or family income, unemployment, not participating in any religious activities, poor living conditions and not having any health insurance plan (Wasti *et al*. 2012:72). Studies also demonstrated variable results on the association of age and sex with adherence behaviour. Some studies revealed that younger age and female sex were significantly associated with low level of adherence (Beer *et al.* 2012:218; Fumaz *et al.* 2008:799; Sullivan *et al.* 2007:3; Wasti *et al. 2012:72*). Individuals taking ART overcome economic obstacles to adherence through the assistance of family and other supporters (e.g. borrowing transport factors, including leveraging social support to overcome financial barriers to care, stabilising routine possibly through additional caregiver support, and disclosing HIV to older

children). Psychosocial factors that are consistently associated with adherence include psychological distress, active drug or alcohol use, social support, knowledge and beliefs about HIV and its treatment (Machtinger & Bangsberg 2006). Depressive symptoms and stress have been correlated significantly with non-adherence (Amberbir *et al.* 2008: Beer *et al.* 2012:218; and Royal *et al.* 2009:452-453). Studies also indicated that the severity of depression and anxiety have a significant influence on adherence to medications. Patients with severe depression were found four times more likely to report lower adherence than patients with minimal depression (Sarna *et al.* 2008:32). A study identified the level of alcohol consumption as a most important alcoholic factor to affect adherence (Parsons *et al.* 2007:365). Another study also demonstrated a significant association between daily alcohol consumption and non-adherence (Protopopescu 2009:602). Smoking and alcoholic habits were also found to be associated with non-adherence to ART (Bonolo *et al.* 2005; Sharma *et al.* 2013:51).

Patient-provider interpersonal dynamics of the physician–patient relationship play an important role in determining a variety of patient outcomes including patient adherence to their treatment recommendations. Cohesive partnerships and effective interpersonal communication, encouragement of participation of patients in their own care, trusting relationship between patients and physicians are important components of patient-provider relationship which promote greater patient satisfaction with medical care, which in turn foster higher levels of adherence (Martin & Di Matteo 2005:192-193). Moreover, the patient's opinion of the provider's competence, the affective tone of the relationship (warmth, openness, and cooperation), and the adequacy of referrals may affect treatment adherence (Machtinger & Bangsberg 2006). Studies indicated a lack of trust in their medical provider was a barrier to adherence (Curioso *et al.* 2010:5). A study was done by De and Dalui (2012:251) also

emphasised the influence of negative attitude of health care providers on adherence behavior of patients (Aragonés *et al.* 2011:19; Watt *et al.* 2010:384).

Aspects of the clinical setting that may influence adherence include access to ongoing primary care, involvement in a dedicated adherence program, availability of transportation and childcare, pleasantness of the clinical environment, convenience in scheduling appointments, perceived confidentiality, and satisfaction with past experiences in the health care system (Machtinger & Bangsberg 2006). In a Meta-analysis of adherence studies issues of access, including financial constraints and a disruption in access to medications were identified as a more common barrier to adherence in developing countries (Mills *et al.* 2006b: 2056). Although access to medication is essential for the continuation of treatment, free access to medication was found to affect adherence negatively. On multivariate analysis patients receiving free ARVs were 4.4 times more likely to report lower adherence than patients paying out-of-pocket for ART (Sarna *et al.* 2008:32).

Patients' follow up status in relation to adherence needs also worth mentioning. The regular clinic follows up was found to be significantly associated with good adherence (Cauldbeck *et al.* 2009). In another study, respondents who missed a clinic appointment were almost three times more likely to report poor adherence, compared with those who never missed a clinic appointment (Watt *et al.* 2010:385). A study by Chalker *et al*. (2010) also revealed attendance of patients within three days of appointment was associated with adherence to ART. Several studies have also evaluated the reasons for non-adherence to antiretroviral therapy. The most commonly cited reasons were forgetfulness, being away from home and being busy with other things (Hansana *et al.* 2013). Transport costs and financial problems were also mentioned in some studies as a reason for non-adherence (Mills *et al.* 2006b:2056; Wasti *et al.* 2012:75). Studies also identified running out of medications (Wasti *et al.* 2012:75; Tiyou *et al.* 2010:4), avoiding side effects (Sullivan *et al.* 2007:3; and felt sick (Amberbir *et al.* 2008 as barriers to adherence.

The challenge of ensuring continuity of care for migrant workers has gained increasing attention. Hellen Bygrave's survey report on Targeted adherence strategies for the provision of cross-border antiretroviral therapy to migrant farm workers in Musina, South Africa. It indicates that among migrant workers moving from South Africa and Zimbabwe, access to antiretroviral treatment (ART) is often denied because of concerns about adherence and continuity of care as less than 50% of those in need of ART were able to access care. Her methodology included using specific counselling tools developed to outline transfer steps and the potential regimen and formulation changes that may occur to maintain their clients on ART. Furthermore, providing patients with a patient-held record and questioning them on their travel plans at each visit (monthly during the first year). She hails the application of aspects of countries adopting a standard first-line ART regimen in a fixed-dose combination and adopting a regular patient-held health record to enhance continuity of care as an end to addressing the patient mobility challenge in all ART programs which could reduce loss to follow-up (Cutsem *et al*. 2012).

In her study on Trends in Loss to Follow-Up among Migrant Workers on Antiretroviral Therapy in a Community Cohort in Lesotho, (Bygrave *et al*. 2010) argues that among the many challenges associated with widespread access to HIV/AIDS care in Lesotho, the provision of treatment and care to mobile populations poses particular challenges. She enlisted the barriers to care for migrants as including access to health services and social support in the host country, discrimination, limited ability to take sick leave and stigma associated with taking medicine at work. The findings of a rapid survey that assessed the rates of loss to follow-up for migrant workers compared with non-migrant workers carried out among a non-random selection of defaulters indicate that the most common reason for defaulting was migrating to South Africa for work. (Bygrave *et al*. 2010).

2.4.2 Designing a data mining and analytics tool to predict possible LTFUs among mobile PLHIVs.

According to a study conducted by Fraser *et al* 2007 their preliminary data demonstrated a number of benefits in tracking patients who have not been prescribed congruous drugs, those who fail to return for follow-up, and those who do not have medications collected for them by health care providers. Supplemental benefits were visually perceived in providing access to key laboratory data analyzed and used to ameliorate the timeliness and quality of care. They highlight that follow-up was achieved by a coalescence of reports from information systems along with CHW teams. They affirm that incipient technologies such as low-cost satellite Internet access, hand-held tablets, and perspicacious mobile phones are availing to expand the spread of these electronic systems (Fraser *et al*. 2007).

Given the well-known barriers that prevent patients from returning for follow-up in resource-poor environments and the frequent need for crossing borders due to poverty and social disruption, the responsibility is placed on the health system to find missing patients and provide treatment. They warn that failure to do so with HIV is often a life-threatening situation. They recommend that an essential component of the information system is a master patient list which should be regularly updated and used to find missing or failing patients. Point worth noting is that this list should include a system for unique patient ID numbers. They give an example of their experience which shows that if the project has more than one treatment site and/or more than a few hundred patients, this master patient list must be computerised and be part of an electronic medical record (EMR) system (Fraser *et al*. 2007).

They conclude that efficacious information systems in developing countries are a recent innovation but require playing an incrementing role in fortifying and monitoring HIV projects as

they scale up from thousands to hundreds of thousands of patients. They note that a particular focus should be placed on tracking patients from initial diagnosis to initiation of efficacious treatment and then monitoring them for treatment breaks or loss to follow-up (Fraser *et al*. 2007).

In rural Rwanda, PIH runs a group of six clinics in and around Rwinkwavu, and data on HIV patients are managed with a newer version of the HIV-EMR based on the OpenMRS architecture. Patient follow-up visits are logged through the entry of follow-up forms, and patients who fail to return are highlighted in monthly reports. This allowed staff to rapidly identify a serious decline in follow-up among patients who had stopped receiving food. An alternative way is to determine if patients are adhering to their regimen is by tracking when community health care workers are returning to collect medications. Several projects have used cell phones to assist in patient follow-up and provide access to medical data such as laboratory tests. PIH in Rwanda sends SMS messages warning of abnormal lab results to clinical staff. Mobile phones have also been tested as a way to encourage patient compliance with TB treatment in South Africa and to monitor medication side effects for treatment of sexually transmitted infections in Peru (Fraser *et al*. 2007).

In summary, designing an efficient electronic tracking tool with the involvement of the users and actors at various reporting levels is key to reducing rates of LTFUs. There are significant barriers to follow-up that include disconnected telephones, changes of address, new employers, or imprisonment. The focus should be on follow-up early in the design of the tool to overcome the barriers they encounter as early as possible.

2.4.3 Developing a data mining and analytics tool to predict possible LTFUs among mobile PLHIVs.

In the study "Strategies to improve patient retention on antiretroviral therapy in sub-Saharan Africa", the researchers advice that every facility that provides ART must set up a simple, standardized monitoring system to track the numbers of patients starting therapy every month or every quarter and to determine at the end of every quarter five key standardized outcomes – those who are alive and on therapy, those who are dead, those who are known to have stopped treatment, those who have transferred out to another facility, and those who have been lost to follow-up or 'defaulted'. Being alive and on therapy implies retention in care. Formal transfer outs from one facility to another are common as the number of facilities expands and patients seek treatment closer to home (Harries *et al*. 2010).

EMRs can play an important role, particularly as patients and treatment sites increase. Experience from a wide range of sites across Africa has shown that for this to work properly, however, adequate human resources and staff training are essential. Harries further explains that the predictable results in resource-constrained settings are poorly completed forms, incomplete data sets, and unreliable data on what counts, namely retention on therapy and attrition. He advises ART programme designers and managers to resist this pressure, as timely collection of reliable data on the five standardised outcomes is hard enough to achieve on its own without over-loading the often manual monitoring system with a host of other parameters (Harries *et al*. 2010).

Patient tracing is the combination of a number of interventions embarked on by a team of HCWs to reach patients who dropped from care. The aim is to encourage them to return and continue on care and support for their own benefits and/or the benefit of the larger population. When developing a system for tracking possible LTFUs, there are a number of tracing steps to be

considered; ability to uniquely enroll/profile mobile patients at the ART/CCC (acquire locator information of the next of kin, village, area, landmarks, area representative, relatives), ability of scheduling appointments days before the appointment based on the enrolment dates, ability of identifying missed appointments from the appointments register, ability to search for and retrieve patient records for editing and updates, ability to schedule notifications to the treatment supporting staff and reminders to the clients on their upcoming appointments. Most importantly is that the tool with up-to-date information should be easily accessible to the health care providers and the CHEWs or the linkage facilitators (The National ART Program Swaziland 2012).

Every patient who does not come on their appointment date should be flagged by the tracking tool, the treatment supporting staff needs to make a follow up at least twice by phone and a third time by a home visit. Each HIV Care Site should have a Focal Person and Back-Up Focal Person who is responsible for overseeing and ensuring timely follow up of patients who have defaulted. This should be the same focal person responsible for ensuring management of the referral system. The team at the HIV Care Site should meet to establish a system for follow- up of all patients. One system could be that the Focal Person does the routine patient tracking when on duty, while another staff assumes one of the Focal Person's other duties. When the Focal Person is not present, the backup Focal Person does the patient tracing. Another system could be a monthly rotation system (coordinated by the Focal Person), whereby a different staff is responsible for patient tracking (The National ART Program Swaziland 2012).

Clinics vary in terms of their ability to trace patients who are either late for their appointment or who have been recorded as lost to follow-up. Well-resourced clinics that are not overly burdened with high patient loads may have outreach teams that can trace patients and reduce their LTFU rates (Harries *et al*. 2010).

2.4.4 Validating and Implementing a data mining and analytics tool for predicting possible LTFUs among mobile PLHIVs.

Measured through accuracy, reliability, and usefulness, validation is the process of assessing how well data mining models perform against real data, it involves understanding the quality and characteristics before deploying it into a production environment. To build a predictive model, the researcher assembles the datasets that used for training. With clear objectives (contains the event or action that the model is to predict and the period when it is likely to happen) formulated, the researcher cleanses and organize the data, perform data treatment including missing values and outlier fixing, make a descriptive analysis of the data with statistical distributions, and create data sets used for the model-building. The Focus was on the behavioural information as it is more powerful for predictions than demographic data. Building the model involves writing model code, building the model using a sample of the data set created, calculating scores, and validating the data (Kalechofsky & Ph 2016).

The researcher used the following approaches iteratively as he created, tested, and refined models to answer to his problems; separate the data into training and testing sets to test the accuracy of predictions, use various measures of statistical validity to determine whether there are problems in the data or in the tool, partitioning data into testing and training sets, filtering models to train and test different combinations of the same source data, measuring lift and gain, a lift chart is a method of visualizing the improvement that you get from using a data mining model, when you compare it to random guessing. Firstly, performing cross-validation of data sets, generating classification matrices. These charts sort good and bad guesses into a table so that you can quickly and easily gauge how accurately the model predicts the target value, creating scatter plots to assess the fit of a regression formula. Secondly, creating profit charts that associate financial gain or costs with the use of a mining model, so that one can assess the value of the recommendations.

Finally, work with health experts to review the results of the data mining tool to determine whether the discovered patterns have a meaning in the targeted health scenario. The model was built on subset of data and when complete and validated, it was run using the PLHIVs data or case base. Assessments included generating rankings from the model to understand the performance of the model. The ranking scores allow a PLHIV to be ranked in order of the predicted score, such as from most likely to least likely to default ARV drugs or from most adherence to least adherence/retention. Consequently, the researcher monitored the performance of the model on an ongoing basis, and take advantage of new data or techniques as they become available (Kalechofsky & Ph 2016).

To measure if the model was successful, the researcher utilized principal success metrics or indicators of model success; compare predictive model performance against random results with lift charts, hold out samples, and decile tables, evaluate the validity of the discovery with target shuffling useful for identifying false positives, or when two events or variables occurring together are perceived to have a cause-and-effect relationship and test predictive model consistency using bootstrap sampling with certain subsets of data over and over again to provide an estimate of accuracy for statistical and operational significance. Implementation efficiency which involves ease of use, immediacy, and practicality of model implementation is important. It affirms if it is costly to deploy a smart model that could be a downside. The time-to-capability (or value) speaks to the best idea in the world having no real business value unless it hits the street and people are using the results. Spending years on a huge abstract idea may be less valuable than quick low-hanging fruit. Under end-user satisfaction; if internal users, sponsors, or end customers are not happy with what they experience, it can be an uphill battle, regardless of how good the model is (Kalechofsky & Ph 2016).

2.5 Gaps in Literature

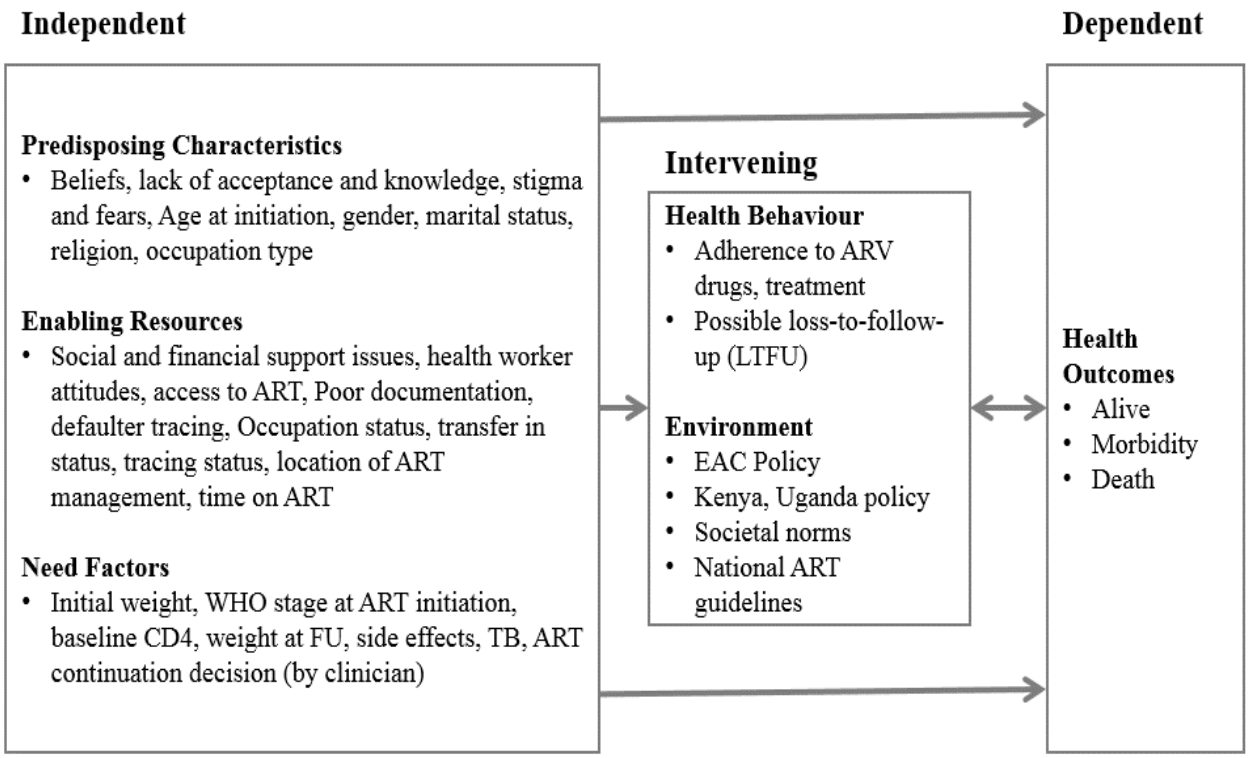
The researcher highlighted literature gaps found while reviewing published documents related to the study area. Strategies such as improving early HIV diagnosis and the early initiation of ART, providing comprehensive nutrition support, implementing intensive monitoring using newly developed indicators, and tracking patients who do not return for their appointments are needed (Oo *et al*. 2016). In their study "Alarming attrition rates among HIV-infected individuals in pre-antiretroviral therapy care in Myanmar, 2011-2014", the researchers propose the need to ensure that the systematic patient tracking systems are utilised for all patients and not limited only to those on ART.

There is evidence that enhanced documentation can affect patient attrition. A few patients have been known to default ARV drugs in one facility, just to end up in another facility as patients. The researchers think it is important to enhance authentic documentation and offer therapeutic records (to some degree) for patients in care and treatment. The acknowledgment is through more prominent coordinated effort among treatment providers in a geographic catchment (Eguzo *et al*. 2015). Retention is additionally higher for programs with peer group support, especially for programs with more than 2000 clients may be more prone to higher LTFU rates. They suggest that it may be important to limit the size of individual treatment programs to under 2000. Leading to the activation of more treatment sites closer to the clients' communities, thus improving tracking. Sites with "active follow-up" models, such as home visits, have shown lower rates of LTFU. They conclude that these proposals are not comprehensive, but only go to show some tactics that are successful. With the rapid infiltration of mobile technology in Africa, the researchers encouraged program managers to explore the use of mobile technology in patient following and retention (Eguzo *et al*. 2015).

According to their study, the researchers recommended that ART service providers and program implementers should make use of their evidence to focus their attention toward the identified predictors. The limitation of their study was that data were not obtained from the LTFU patients themselves, as tracing the LTFU patients was not easy. Therefore, they proposed that future studies should attempt to reach the LTFU patients to make the evidence more complete (Megerso *et al*. 2016).

2.6 Conceptual Framework

FIGURE 2.1
Conceptual model for possible LTFU on ART



2.7 Operationalization of variables

2.7.1 Independent Variables

These are the factors that influence adherence to ART and care drugs, categorised as;

- **Predisposing Characteristics**; Beliefs, lack of acceptance and knowledge, stigma and fears, Age at initiation, gender, marital status, religion, occupation type
- **Enabling Resources**; Social and financial support issues, health worker attitudes, access to ART, Poor documentation, defaulter tracing, Occupation status, transfer in status, tracing status, location of ART management, time on ART
- **Need Factors**; Initial weight, WHO stage at ART initiation, baseline CD4, weight at follow-up, side effects, Tuberculosis, ART continuation decision.

2.7.2 Intervening Factors

Intervening or mediating variables are hypothetical variables which cannot be observed in an experiment but used to explain causal links between other variables. They do not cause the outcome but may modify it. In this study they are broadly categorized into;

- **Health behavioral factors** (service appointment, adherence patterns and missed appointments/defaulting ARV drugs).
- **Environmental factors** (regional, national and district policies, societal norms, national ART guidelines).

2.7.2 Dependent Variables

These are the factors being affected by the independent variables, adherence of PLHIVs to ART and therefore determining their morbidity (improved immunity), staying alive longer or death.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

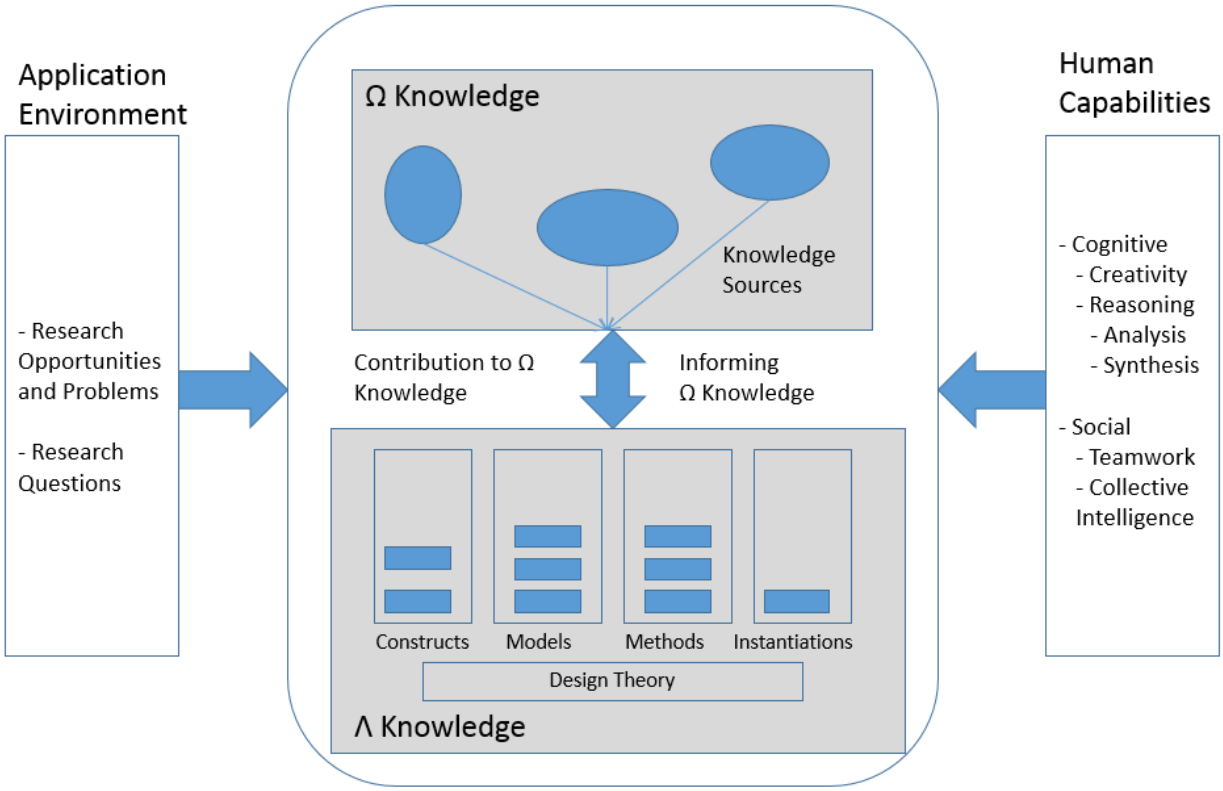
This chapter presents the methodology that used in this research. The chapter, therefore, discusses research design, site of study, target population, sampling techniques and sample size, research instruments, validity and reliability of instruments, pilot study, data collection procedures, data analysis, presentation, data management and ethical consideration.

3.2 Research Design

This study implemented a Design Science (DS) methodology to develop a tool as an embedded solution to this research problem, design science is solution-oriented, and elicits knowledge in the form of constructs, techniques and methods, models, well-developed theory for mapping—the know-how for creating artifacts that satisfy given sets of functional requirements. It involves a rigorous process to design artifacts (*models, methods, instantiations, social innovations or new properties of technical, social, and/or informational resources*) to solve observed problems, to make research contributions, to evaluate the designs, and to communicate the results to appropriate audiences. Peffers notes that design science research must produce *an artifact created to address a problem*. Further, the artifact should be relevant to the solution of an unsolved and important business problem. He further explains that the solution's *utility, quality, and efficacy* must be rigorously evaluated and represent a verifiable contribution, rigor must be applied in both the development of the artifact and its evaluation. He concludes that, the development of the artifact should be a search process that draws from existing theories and knowledge to come up with a solution to a defined problem and that the research must be effectively communicated to appropriate audiences (Peffers *et al.* 2009).

Alan R. Hevner from the *University of South Florida* defines design Science as a creative research paradigm that informs multiple audiences. He defines the design process as a *sequence of expert activities producing an innovative product*, he emphasizes that the artifact enables the researcher to get a better grasp of the problem and that the re-evaluation of the problem improves the quality of the design process. He explains that the build-and-evaluate loop is typically iterated a number of times before the final design artifact is generated.

FIGURE 4
Design Science Process



He notes that in design science, the focus is on the so-called field-tested and grounded technological rule as a possible product of research with the potential to improve the relevance of

academic research in management. He *presents seven principles and guidelines* of design science required and that should be adhered to for the success of the study.

TABLE 2
Design Science Principles and Guidelines

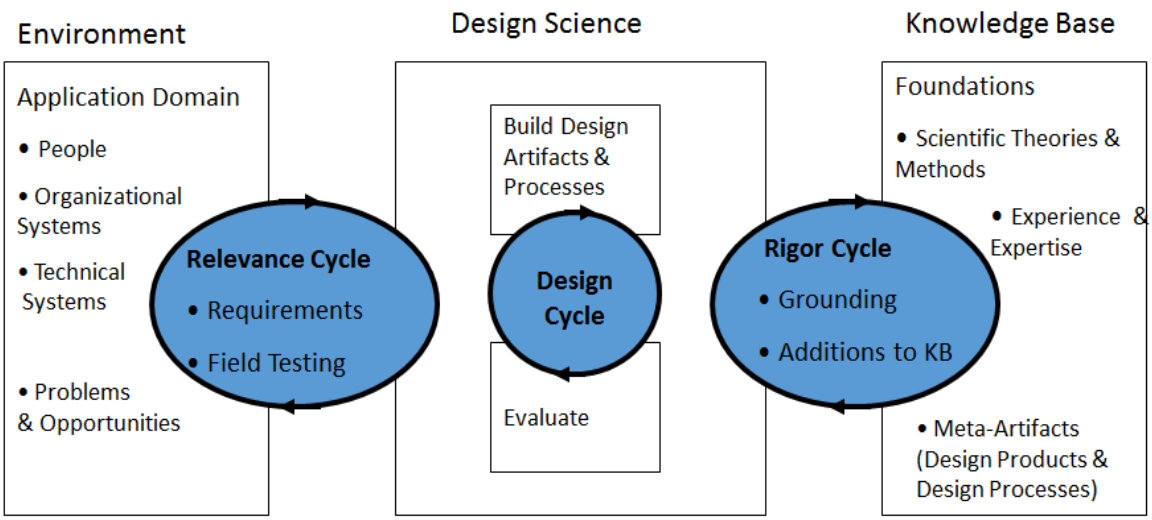
Guideline/Principle	**Description**	**Study Output**
Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.	Data Mining model for predicting possible LTFUs
Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.	Tool with data mining model for predicting possible LTFUs
Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.	Evaluation and implementation schedules of the tool after development
Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.	The absence and introduction of the tool with demonstration of improved adherence through tracking
Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.	Users, clients and developers involved in development process of the tool
Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.	Existing systems are being reviewed and gaps used to inform design of the tool
Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.	Tool was implemented at participating facilities with stake holders and technical data teams involved

In summary, design science requires the creation of an innovative, purposeful artifact for a special problem domain. The artifact must be evaluated in order to ensure its utility for the specified problem. In order to form a novel research contribution, the artifact must either solve a problem that has not yet been solved, or provide a more effective solution. Both the construction

and evaluation of the artifact must be done rigorously, and the results of the research presented effectively both to technology-oriented and management-oriented audiences.

A Hevner in his article "*A Three Cycle View of Design Science Research, Scandinavian Journal of Information Systems*" explains that design science can be seen as an embodiment of three closely related cycles of activities where the relevance cycle initiates design science research with an application context that not only provides the requirements for the research as inputs but also defines acceptance criteria for the ultimate evaluation of the research results. The rigor cycle provides past knowledge to the research project to ensure its innovation. The central Design Cycle iterates between the core activities of building and evaluating the design artifacts and processes of the research.

FIGURE 5
Three cycles of Design Science



He concludes that, it is dependent on the researchers to thoroughly research and reference the knowledge base in order to guarantee that the designs produced are research contributions and not routine designs based upon the application of well-known processes. This methodology was

extensively and exhaustively used in the achievement of the study's specific objectives two, three and four on design, development, implementation and validation of the tool.

Profile information with PLHIV data gathered using the available enrollment and ART daily activity registers was sufficient in generating data for main factors achieved in specific objective one. Eric Biller in his article "*how to design a mobile app using user interface design principles*" states that if you want to design a great, functional mobile app interface, design principles are hugely important. Design principles commonly called heuristics are loose guidelines that designers can stick to in order to improve the quality of a user interface design. He highlights that it is important to ensure that one understands the main tasks they want to accomplish on their app through research and make those the focus of the interface (Biller 2016).

In designing the tool, the researcher utilized the six principles of user interface design listed below;

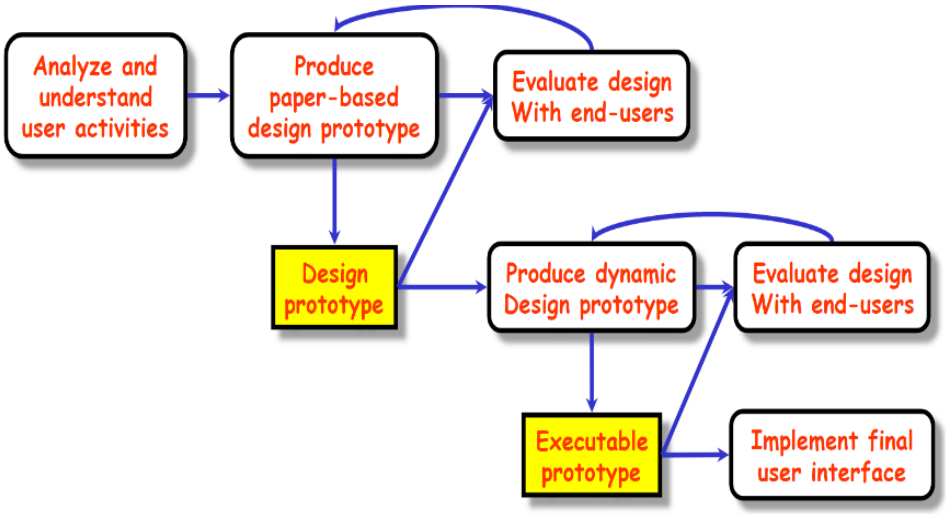
1. **User familiarity** - The interface should use terms and concepts which are drawn from the experience of the people making most use of the system.
2. **Consistency** - The interface should be consistent in that, wherever possible, comparable operations should be activated in the same way
3. **Minimal surprise** - Users should never be surprised by the behavior of a system.
4. **Recoverability** - The interface should include mechanisms to allow users to recover from possible errors.
5. **User guidance** - The interface should provide meaningful feedback when errors occur and provide context-sensitive user help facilities.
6. **User diversity** - The interface should provide appropriate interaction facilities for different types of system user

There are main usability attributes to be considered in developing the UI of an application, they include;

1. **Learnability**: How long does it take a new user to become productive with the system?
2. **Speed of operation**: How well does the system response match the user's work practice?
3. **Robustness**: How tolerant is the system of user error?
4. **Recoverability**: How good is the system at recovering from user error?
5. **Adaptability**: How closely is the system tied to a single model of work?

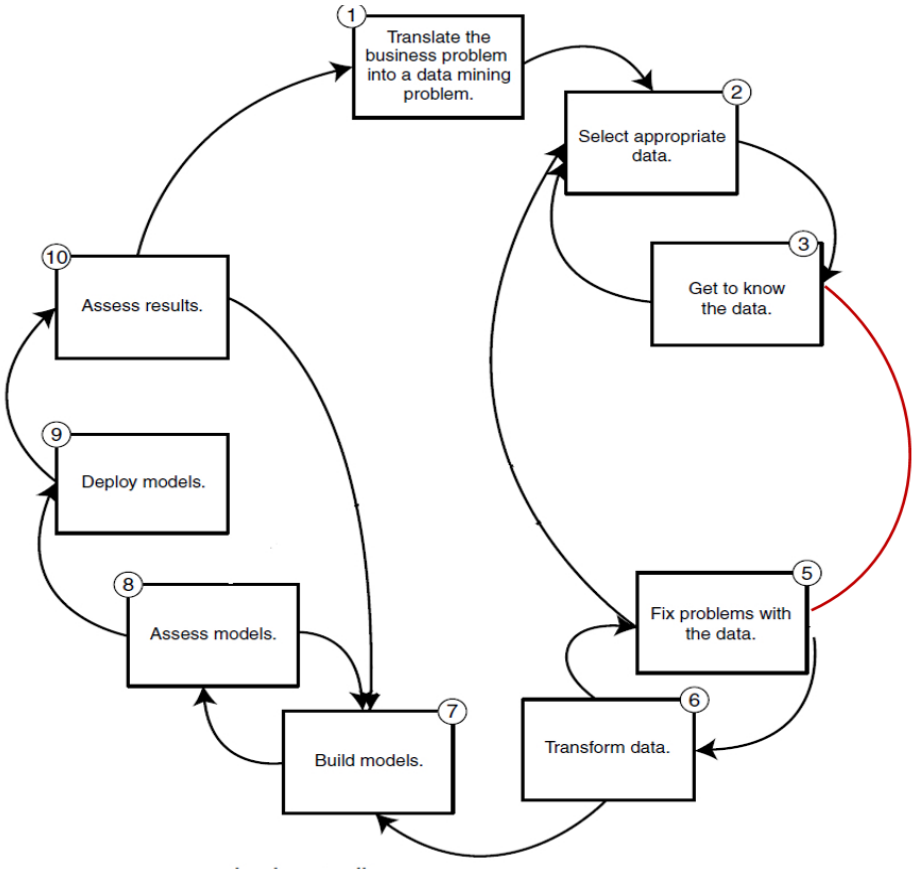
Interface design process involves an eight step flow of interrelated tasks carried out routinely; the analysis and understanding of user interactions with the application, production of paper based draft prototypes of the application, designing the prototype and evaluating it iteratively with end-users. The developer then produces a dynamic design prototype that is then transformed into an executable prototype and its design evaluated by end users through tests done, an evaluation check-list and review documentation, the results of the evaluation would inform the implementation of the final user

FIGURE 6
User interface design cycle



It is clearly stated in Berry and Linnof's literature that "Building a predictive model requires separation in time between the model inputs or predictors and the model output, the thing to be predicted. If this separation is not maintained, the model does not work" (Berry & Linoff, 2009 – Data mining techniques, 2nd Edition). The data mining process is best thought of as a set of nested loops rather than a straight line, the steps have a natural order, but it is not necessary or even desirable to completely finish with one before moving on to the next. In most cases, things learned in later steps cause earlier ones to be revisited. Through the adapted the Berry and Linnof's data mining methodology with additions on the removal of the "Create a model set" process and linking "get to know data" to "fix problems with data" (Beatrice & Mwendia 2017).

FIGURE 7
Berry and Linnof's Data Mining Methodology



3.3 Target Population

The target population are PLHIVs at the Busia border while the sample population was PLHIVs at the Busia border profiled as mobile. Based on the projected population of Busia, Kenya: 905,658 Busia, Uganda: 319,247 and the HIV prevalence of 6.8%, 7.1% respectively, the calculated population of persons living with HIV/AIDS in Busia site is approximately 65,530. However, this study targets adult clients enrolled in care and treatment at the CCC in Busia County Referral Hospital and ART clinic at Busia Health Center IV (Red Cross).

TABLE 3
Target Population

Target population	**Study Site and Coverage**	**Participating Facilities**
• People Living with HIV/AIDS (PLHIV)	• Five (5) kilometre radius from the border crossing point and along the major transport corridor of Busia county in Kenya and Busia District in Uganda	• Busia County Referral Hospital (BCRH) on the Kenyan side of the border • Busia Health Centre IV (Red Cross) on the Ugandan side of the border
Sample Population	**Criteria for selection**	**Inclusion/Exclusion Criteria**
• People Living with HIV/AIDS (PLHIV) profiled as mobile	• An individual who receives HIV counselling is tested and confirmed to be Positive • People who self-identify as HIV-positive • Enrolled in care and treatment or on ART at the participating facilities • Reports occasional or regular movement from their usual residence including cross-border	• Men and women over the age of 18 years • Anyone under the age of 18 was be excluded • There is no exclusion based on race, gender or ethnicity, and pregnant women were not excluded. • Respondents were excluded if they are unable to understand the study and informed consent process (for example, if they are intoxicated or do not comprehend the conversation).

3.4 Sampling and Sampling procedure

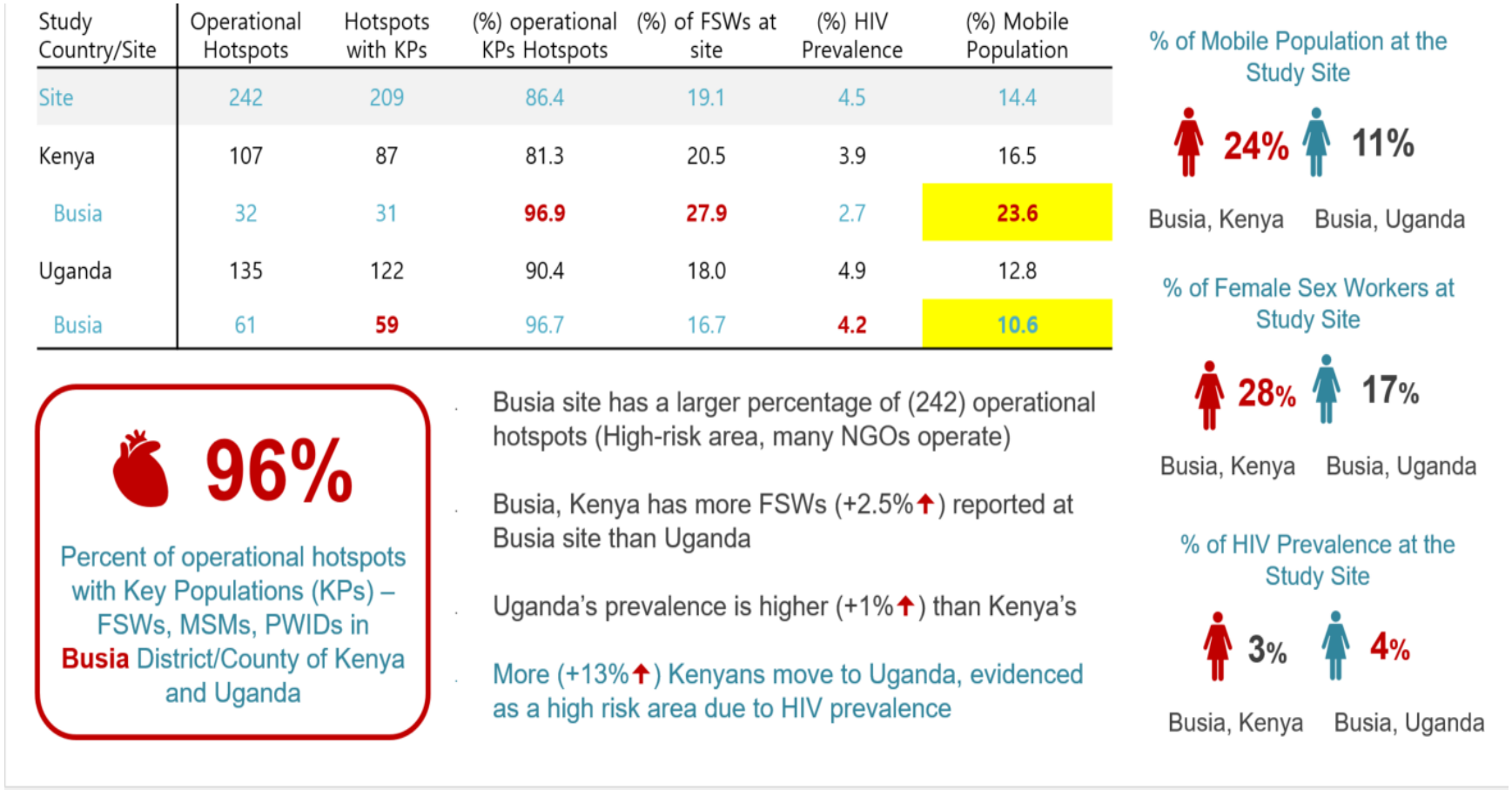
A probabilistic, stratified sampling technique was used to select the PLHIVs at the site, a purposive sampling technique was used for selecting the Health Care Providers and a simple random technique was used for the CHEWs/CHWs and VHTs/linkage facilitators. Sample size has been calculated for each target people in each participating site within Kenya/Uganda based on the available information about the outcome variable (end point: possibility of becoming LTFU in the case of PLHIV). Sample size calculation is based on the proportion testing HIV positive and linked to care in each participating country. In the case of PLHIV**,** the calculated sample is compared to the number of PLHIV registered in the facilities within the project catchment area (less than 100 each). Data from each site of the border was combined with the achieved sample on the opposite side of the border for cross-border analysis in Busia, Kenya/Uganda. Sample size calculations are based on the estimation of a single proportion with a 95% confidence interval (CI) for the desired precision level. Sample sizes for each population sub-group included in the assessment have been calculated the single proportional formula was used to calculate the sample size for each target population:

$$\mathbf{N} = \frac{\mathbf{z^2 p\,(1-p)\ x\ D}}{\mathbf{d^2}} \quad (1)$$

Where by

N	=	Sample size
z	=	Standard normal value for 95% confidence
P	=	proportion of the target population accessing health services
d	=	margin of error (5%)
D	=	design effect (for sampled population using RDS or cluster sampling)

FIGURE 8
Measure Evaluation Assessments on cross-border sites (July, 2017)



Study Country/Site	Operational Hotspots	Hotspots with KPs	(%) operational KPs Hotspots	(%) of FSWs at site	(%) HIV Prevalence	(%) Mobile Population
Site	242	209	86.4	19.1	4.5	14.4
Kenya	107	87	81.3	20.5	3.9	16.5
Busia	32	31	**96.9**	**27.9**	2.7	**23.6**
Uganda	135	122	90.4	18.0	4.9	12.8
Busia	61	**59**	96.7	16.7	**4.2**	**10.6**

Measure Evaluation conducted an assessment at cross-border sites prioritized and labeled by the EAC and IGAD as those that have high positivity yields. They have key populations and higher number of hotspots and dens where risky sexual behaviors take place.

TABLE 4
Sample Size

Country	Expected baseline value (p)	Margin of error (d)	Required sample size (N)	Adjustment for refusal (non-response – 5%)	Adjustment for unable to contact – 30%	% Predicted Mobile Population	Required mobile sample size (N)
Busia, Kenya	94%	5%	88	92	120	23.63	28
Busia, Uganda	94%	5%	121	127	165	10.56	17
Busia Site	94%	5%	209	219	285	17.10	46

In this study, 5% was added to the sample size to compensate for persons that the researcher is unable to contact and an additional 30% to compensate for non-response.

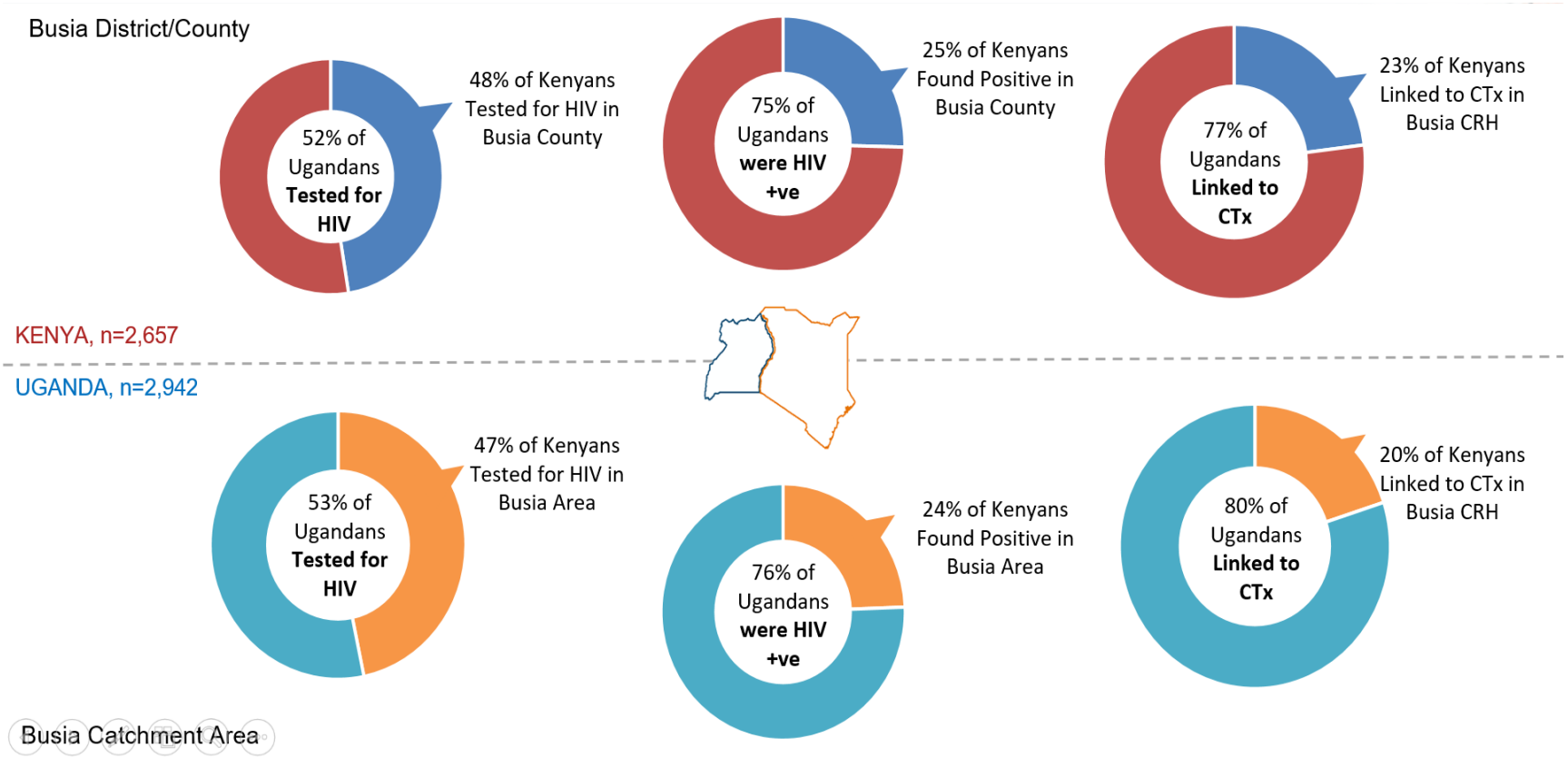
3.5 Available PLHIV Data

The Cross-Border Health Integrated Partnership Project (CB-HIPP) implements activities in Busia District and county, consequently it conducts Integrated Health Service Outreaches with involvement of HCWs from two Health Facilities, and the project conducted an HTS outreach in Busia County/District of Kenya/Uganda in July 2017 reaching 5,599 individuals with HIV counseling and testing services who also received their results. Those found to be HIV positive were linked to care and treatment immediately through the "Test and treat model"

FIGURE 9
Findings of the Integrated Health Service Outreach by CB-HIPP

Country (N=5,599)	**Number of Individuals reached with HIV counseling and testing services in Busia County/District**								
Indicator	# Tested for HIV/AIDS			# found HIV Positive: n (%)			# linked to Treatment: n (%)		
Gender	T	M	F	T	F	M	T	F	M
Kenya	2,657	1,092	1,565	46(1.7)	24(2.2)	22(1.4)	29(63)	17(70.8)	12(54.4)
Uganda	2,942	1,422	1,520	135(**4.6**)	79(**5.6**)	56(3.7)	97(**71.9**)	50(63.3)	47(**83.9**)
District/County	**5,599**	2,514	**3,085**	181(3.2)	103(4.1)	78(2.5)	126(69.6)	467(65)	59(75.6)
Busia (n=1,814)	**Number of Individuals reached with HIV counseling and testing services in Busia catchment area**								
Kenya	848	280	568	20(2.4)	13(4.6)	7(1.2)	15(75)	8(61.5)	7(100)
Uganda	966	425	541	62(6.4)	31(**7.3**)	31(5.7)	**60(96.8)**	**30(96.8)**	**30(96.8)**
Catchment Area	**1,814**	705	**1,109**	82(4.5)	44(6.2)	38(3.4)	75(91.5)	38(86.4)	37(97.4)
Percent (n/N), %	**Percent (%) of Individuals reached with HIV counseling and testing services in Busia catchment area**								
Kenya	32%	26%	36%	43%	54%	32%	52%	47%	58%
Uganda	33%	30%	36%	46%	39%	55%	**62%**	**60%**	**64%**
Overall	32%	28%	36%	**45%**	43%	**49%**	60%	57%	63%

FIGURE 10
Highlights from the IHSO



The data shows that generally more Ugandans than Kenyans are being tested for HIV/AIDS and over 50% of those found to be positive were Ugandans in comparison to 25% Kenyans. About 20% of the HIV positives were linked to care and treatment at the cross-border facilities.

FIGURE 11
Kenya - Mobilization, Testing, Enrolment, Care, and Treatment and follow up

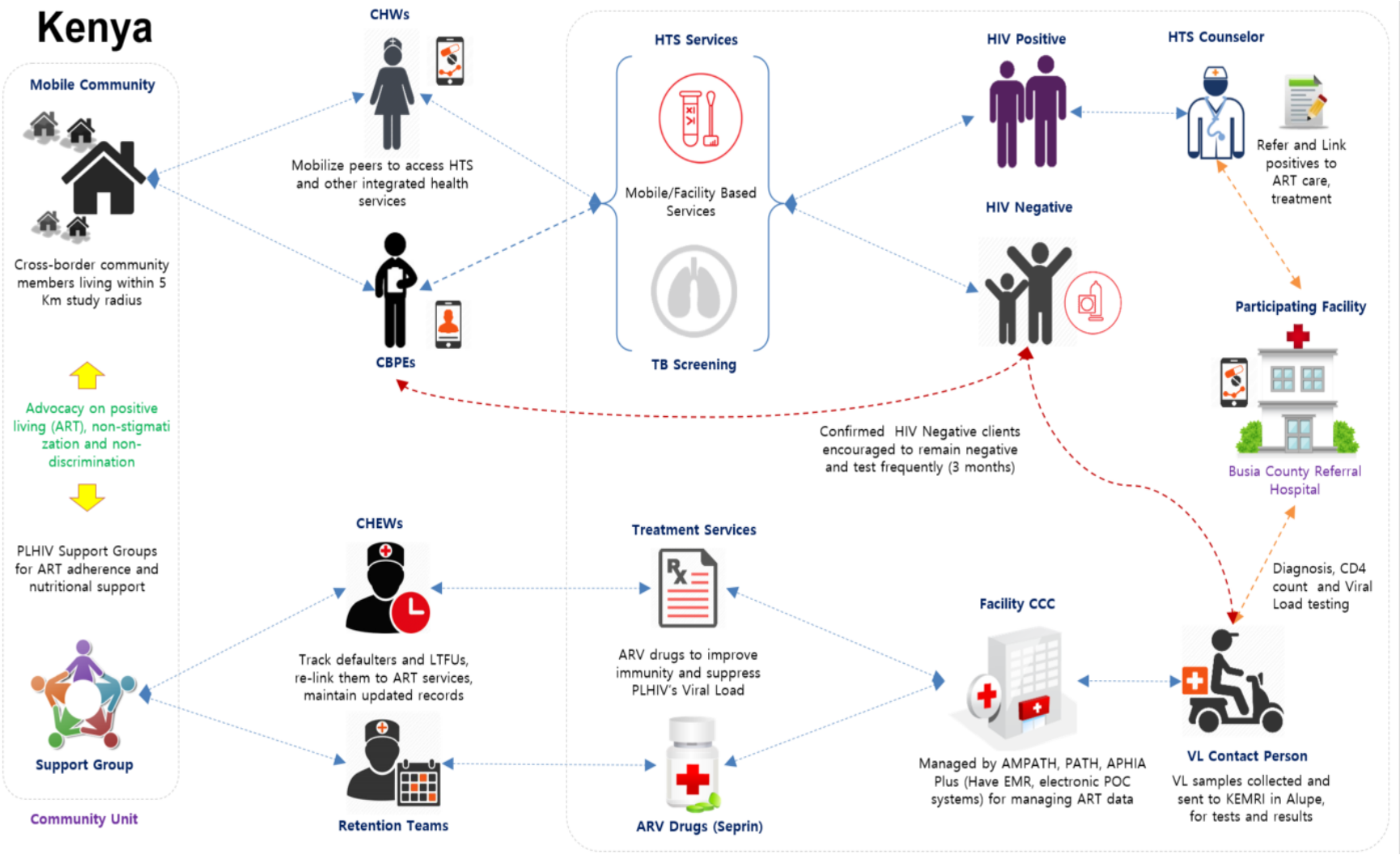


FIGURE 12

Uganda - Mobilization, Testing, Enrolment, Care, and Treatment and follow up

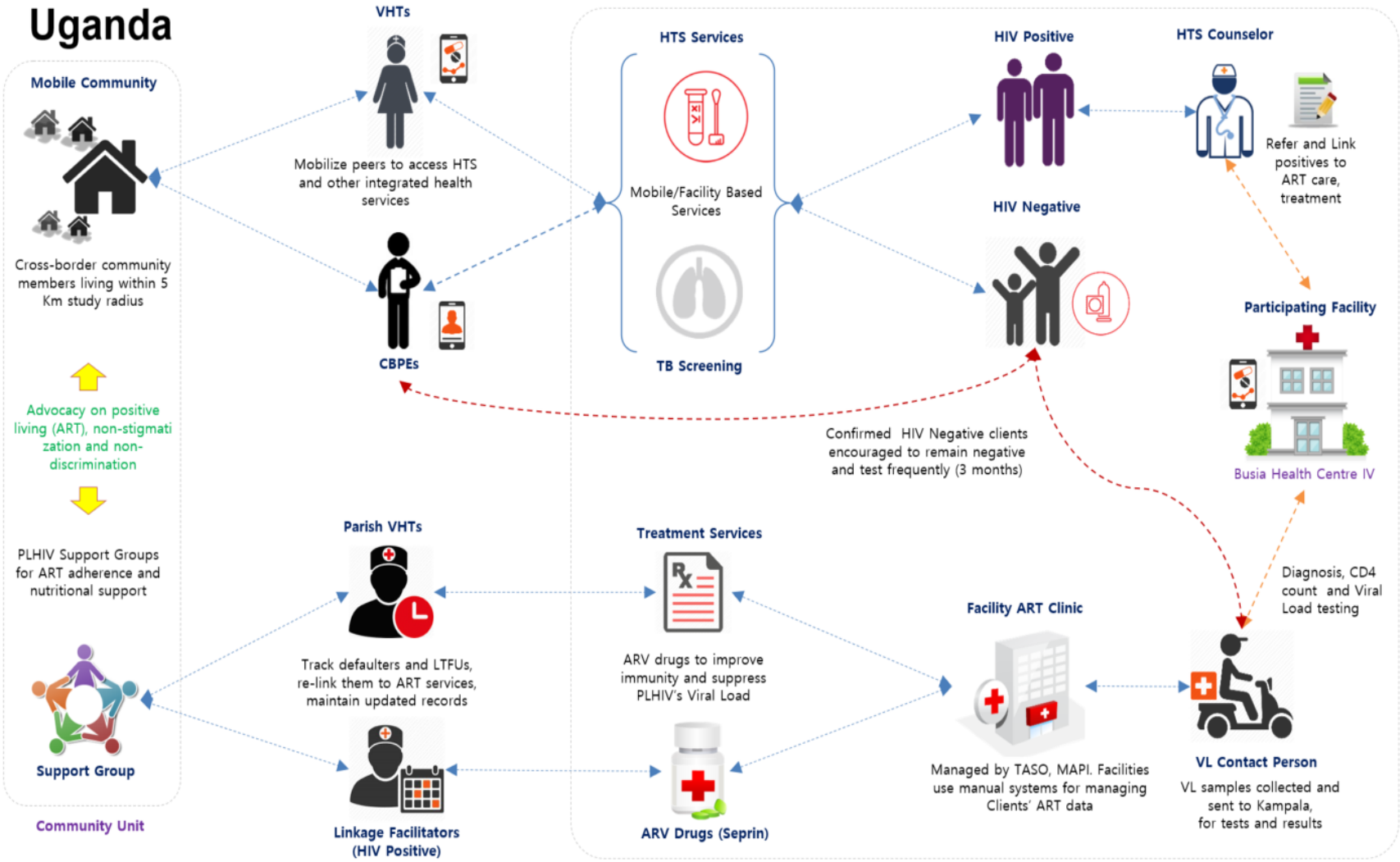


FIGURE 13
Proposed Tool for predicting LTFUs

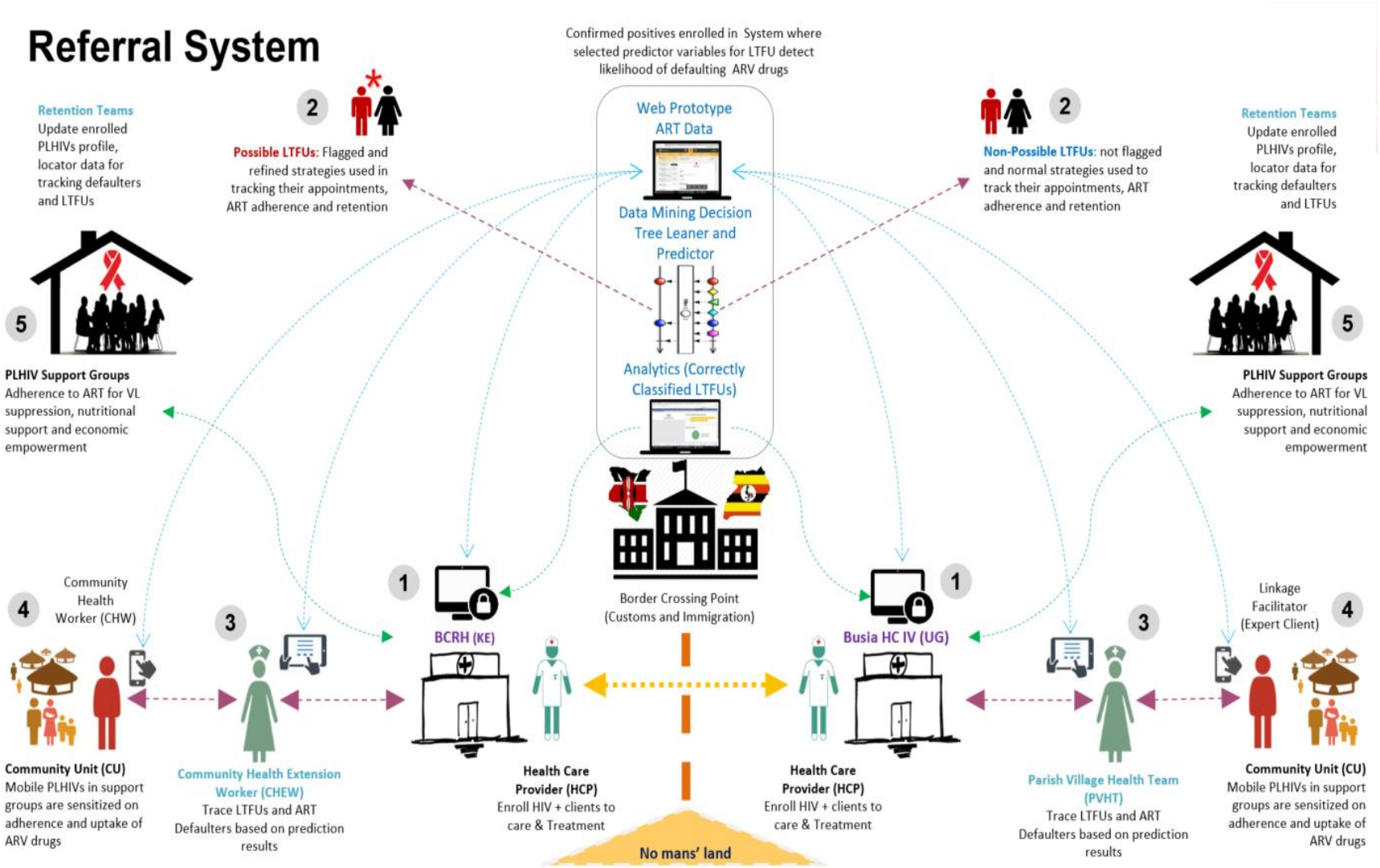


FIGURE 14
Current HIV Algorithm (Kenya/Uganda)

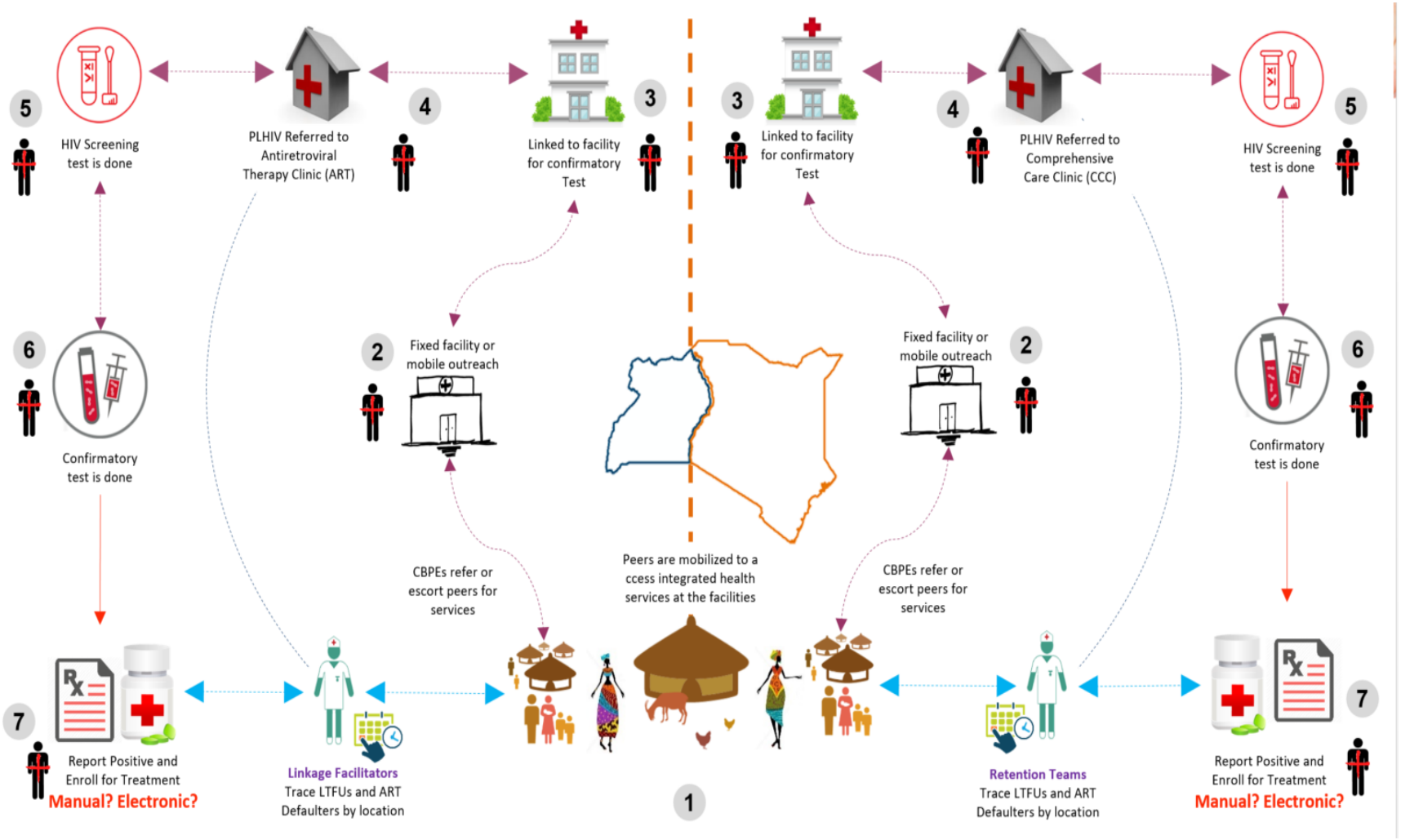
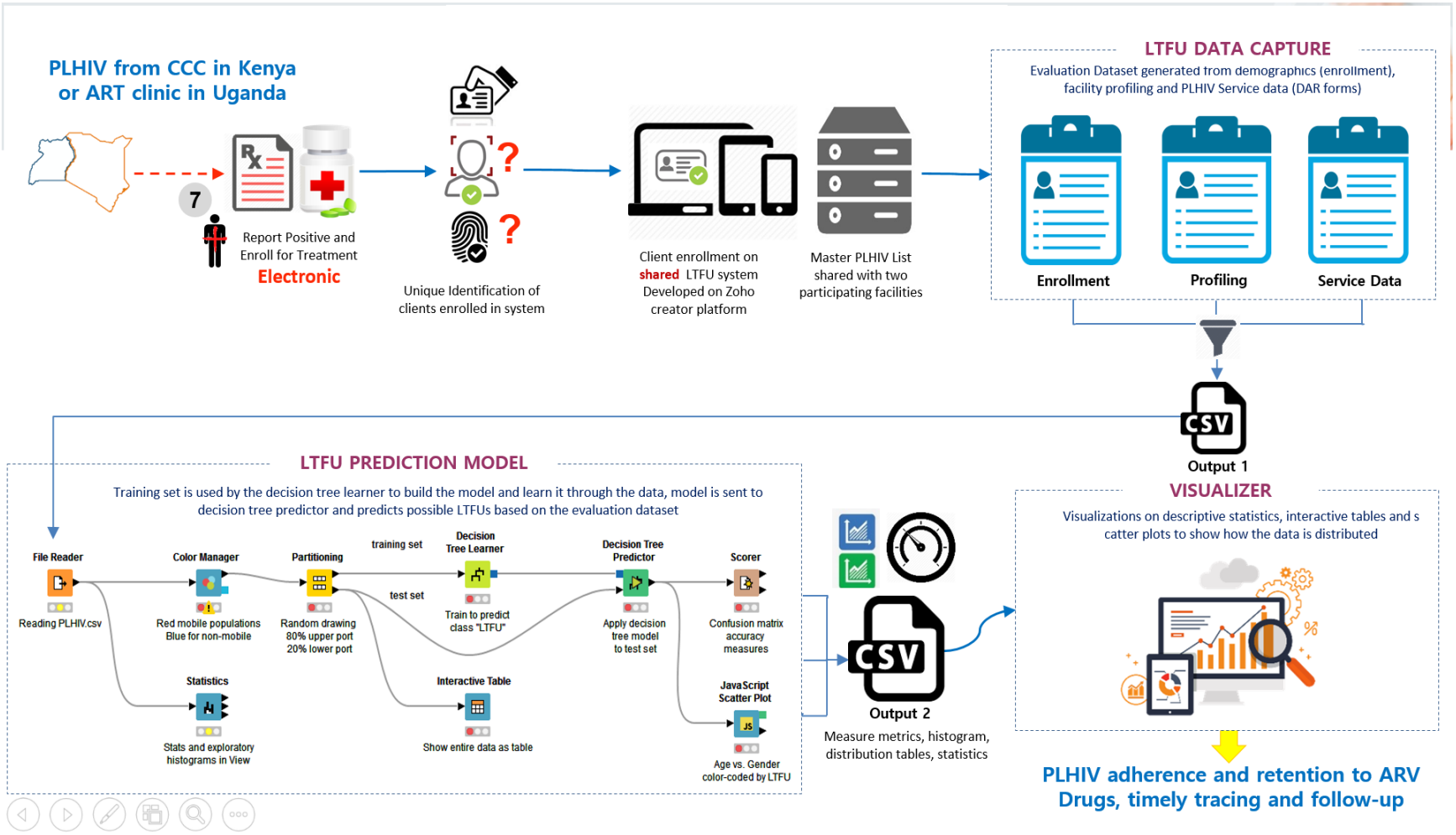


FIGURE 15
Loss-To-Follow-Up Prototype Design



3.6 Data Mining Process

Confidentiality of PLHIV data shared is of great importance. It was de-identified (identification information such as names and phone numbers removed).

Step 1: Translating the business problem into a data mining problem - identifying the problem statement and describing how the problem can be addressed using data mining technics, through review of literature.

- **Step 2: Select the appropriate data -** the required PLHIV enrollment and service data is resident at the participating facilities, cleansed, available, historically accurate, and frequently updated during client's visits. Datasets are selected from the Kenya and Uganda facilities. A training dataset of 45 cases (17 - Uganda, 15 – Kenya together with an additional 13 who were followed up and are currently linked hence 42 not linked: 181-139) was gathered from two participating facilities in liaison with the C/DHMTs, facility in-charges and the CCC/ART clinics coordinators.

- **Step 3: Get to Know the Data –** there is importance in spending time exploring the data before rushing into building models. Along the way, you are likely to discover many data quality problems and be inspired to ask many questions that would not otherwise have come up. Through Computer Human Inspection (CHI), datasets acquired from BCRH, Kenya and BHCIV, Uganda on clients found to be HIV positive during outreach conducted at the sites in July. Most variables have similarities (patient demographics, pretest, tested as HIV results and where referred. Separate data sources (HTS and DAR) linked together using Unique client/Patient ID and demographic details

FIGURE 165
Busia County Referral Hospital Dataset - Kenya

	site	facility	nationality	telephone	age	age_cat	gender	village_hotspot	residence	marital	tested_as	HIV test	Kit name	HIV test 2	Kit name	HIV test 3	Final_result	Refer to	HTC provider
84	BUSIA	Busia CRH	Ugandan	782506136	37	25-49 Yrs	Female	BURUMBA	No	Single	Individual	Positive	38J2916D	N/A	D.F.R	NA	Positive	NA	JAMES
102	BUSIA	Busia CRH	Kenyan	712160375	48	25-49 Yrs	Male	MARACHI B	Yes	Maried Monogamous	Individual	Positive	38J2916D	Positive	D.F.R	NA	Positive	OGC	JUDY
108	BUSIA	Busia CRH	Kenyan	711562844	40	25-49 Yrs	Female	ALUPE	Yes	Single	Individual	Positive	38J2916D	Positive	D.F.R	NA	Positive	OGC	JUDY
123	BUSIA	Busia CRH	Ugandan	702319158	19	15-24 Yrs	Female	MARACHI	Yes	Single	Individual	Positive	38J2916D	Positive	D.F.R	NA	Positive	CCC	MARGRET
142	BUSIA	Busia CRH	Ugandan	773406054	24	15-24 Yrs	Female	SOFIA	Yes	Maried Monogamous	Individual	Positive	DT.FR	Positive	D.F.R	Positive	Positive	NA	SOPHY
155	BUSIA	Busia CRH	Kenyan		53	50+ Yrs	Female	MARACHI	No	Maried Monogamous	Individual	Positive	DT.FR	Positive	D.F.R	NA	Positive	NA	ESTHER
232	BUSIA	Busia CRH	Ugandan	787459506	29	25-49 Yrs	Female	MAWERO	Yes	Divorced	Individual	Positive	DT.FR	Positive	DT.FR	NA	Positive	CCC	LINUS
347	Busia	Busia CRH	Kenyan	708105610	44	25-49 Yrs	Male	Burumba	Yes	Divorced	Individual	Positive	DT.FR	Positive	DT.FR	Positive	Positive	CCC	Teresia
379	Busia	Busia CRH	Kenyan		30	25-49 Yrs	Male	Marachi	Yes	Maried Monogamous	Individual	Positive	DT.FR	Positive	DT.FR	Positive	Positive	CCC	Fred
425	BUSIA	Busia CRH	Kenyan	722486961	35	25-49 Yrs	Female	TESO	Yes	Divorced	Individual	Positive	D.F.R	Positive	D.F.R	NA	Positive	CCC	GETRUDE
580	BUSIA	Busia CRH	Kenyan	N/A	22	15-24 Yrs	Female	BURUMBA	Yes	Maried Monogamous	Individual	Positive	D.F.R	Positive	D.F.R	NA	Positive	CCC	LINUS
613	BUSIA	Busia CRH	Kenyan	770818560	30	25-49 Yrs	Male		Yes	Single	Individual	Positive	D.F.R	Positive	D.F.R	NA	Positive	CCC	PAMELLA
628	BUSIA	Busia CRH	Kenyan	797049011	45	25-49 Yrs	Female	MAUKO	Yes	Divorced	Individual	Positive	P	Positive	D.F.R	NA	Positive	CCC	GETRUDE
636	BUSIA	Busia CRH	Kenyan	791165755	23	15-24 Yrs	Female	SOFIA	No	Divorced	Individual	Positive	P	Positive	D.F.R	NA	Positive	CCC	GETRUDE
714	BUSIA	Busia CRH	Ugandan	725158221	22	15-24 Yrs	Female	SOFIA	No	Married Polygamous	Individual	Positive	D.F.R	Positive	D.F.R	NA	Positive	CCC	GETRUDE
717	BUSIA	Busia CRH	Kenyan	700182278	20	15-24 Yrs	Female	TESO	Yes	Maried Monogamous	Individual	Positive	D.F.R	Positive	D.F.R	NA	Positive	CCC	GETRUDE
725	BUSIA	Busia CRH	Kenyan	N/A	67	50+ Yrs	Male	BURUMBA	Yes	Married Polygamous	Individual	Positive	D.F.R	Positive	D.F.R	NA	Positive	NA	GETRUDE
743	BUSIA	Busia CRH	Ugandan	798530333	30	25-49 Yrs	Female	TOWNSHIP	Yes	Divorced	Individual	Positive	D.F.R	Positive	D.F.R	NA	Positive	CCC	TERESIA
792	BUSIA	Busia CRH	Kenyan	707445492	45	25-49 Yrs	Male	MARACH	Yes	Maried Monogamous	Individual	Positive	D.F.R	Positive	D.F.R	NA	Positive	CCC	FRED
829	BUSIA	Busia CRH	Kenyan	726260546	47	25-49 Yrs	Male	KARIBUNI	Yes	Maried Monogamous	Individual	Positive	D.F.R	Positive	D.F.R	Positive	Positive	CCC	MARGRET

FIGURE 176
Busia Health Centre IV Dataset - Uganda

	country	site	facility	HSU	Nationality	contact	age	gender	residence	xborder	xcounty	marital	Village/Zone/Cell	firsttimetested	tested_more_twice	pretest	tested	got_results	results_HIV	tested_couple	results_couple	Discordant results	HTC entry point	linked	Where?
10	Uganda	Busia	BHC IV	Outreach	Kenyan	792346380	18	Female	No	1	5	Never Married	CUSTOMS	Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
26	Uganda	Busia	BHC IV	Outreach	Kenyan	725554383	22	Male	Yes	1	1	Never Married	CUSTOMS	Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	kenya
59	Uganda	Busia	BHC IV	Outreach	Kenyan	723808084	23	Male	Yes	3	4	Married	SOFIA	No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
65	Uganda	Busia	BHC IV	Outreach	Ugandan	779809401	24	Female	No	1	4	Married	SOFIA	No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
102	Uganda	Busia	BHC IV	Outreach	Ugandan	773326440	45	Male	Yes	2	3	Married	MARACHI	No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
152	Uganda	Busia	BHC IV	Outreach	Ugandan		41	Male	Yes	2	2	Married	MARACHI	No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	YES	KENYA
326	Uganda	Busia	BHC IV	Outreach	Ugandan	782692796	37	Male	Yes	1	2	Never Married	MARACHI	Yes	Yes	Yes	Yes	Yes	positive	No	No	No	No	yes	BUSIA HC IV
335	Uganda	Busia	BHC IV	Outreach	Ugandan		26	Female	Yes	5	2	Never Married	MARACHI	No	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	kenya
342	Uganda	Busia	BHC IV	Outreach	Ugandan	782593215	36	Male	Yes	2	2		SOLO	Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	kenya
346	Uganda	Busia	BHC IV	Outreach	Ugandan	780371133	39	Female	Yes	1	3	Never Married	CUSTOMS	Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
352	Uganda	Busia	BHC IV	Outreach	Ugandan	728348915	23	Male	Yes	1	3	Never Married	CUSTOMS	Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	KENYA
355	Uganda	Busia	BHC IV	Outreach	Ugandan	753112438	22	Male	Yes	1	5		CUSTOMS	No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
357	Uganda	Busia	BHC IV	Outreach	Ugandan	792357119	26	Female	Yes	1	5		CUSTOMS	No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	kenya
363	Uganda	Busia	BHC IV	Outreach	Ugandan		28	Female	Yes	1	5		CUSTOMS	Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
366	Uganda	Busia	BHC IV	Outreach	Ugandan	785357921	29	Male	Yes	1	5	Married		Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
370	Uganda	Busia	BHC IV	Outreach	Ugandan	793402011	33	Male	Yes	1	3	Married		No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
373	Uganda	Busia	BHC IV	Outreach	Ugandan	751722145	47	Male	Yes	1	5	Married		Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
376	Uganda	Busia	BHC IV	Outreach	Ugandan	75751895	37	Female	Yes	1	5	Never Married		No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	BUSIA HC IV
386	Uganda	Busia	BHC IV	Outreach	Ugandan	703999671	17	Female	Yes	1	5	Never Married	CUSTOMS	Yes	No	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	kenya
393	Uganda	Busia	BHC IV	Outreach	Ugandan	76317891	29	Male	Yes	1	5	Never Married	CUSTOMS	No	Yes	Yes	Yes	Yes	positive	No	No	No	C. outreach	yes	kenya

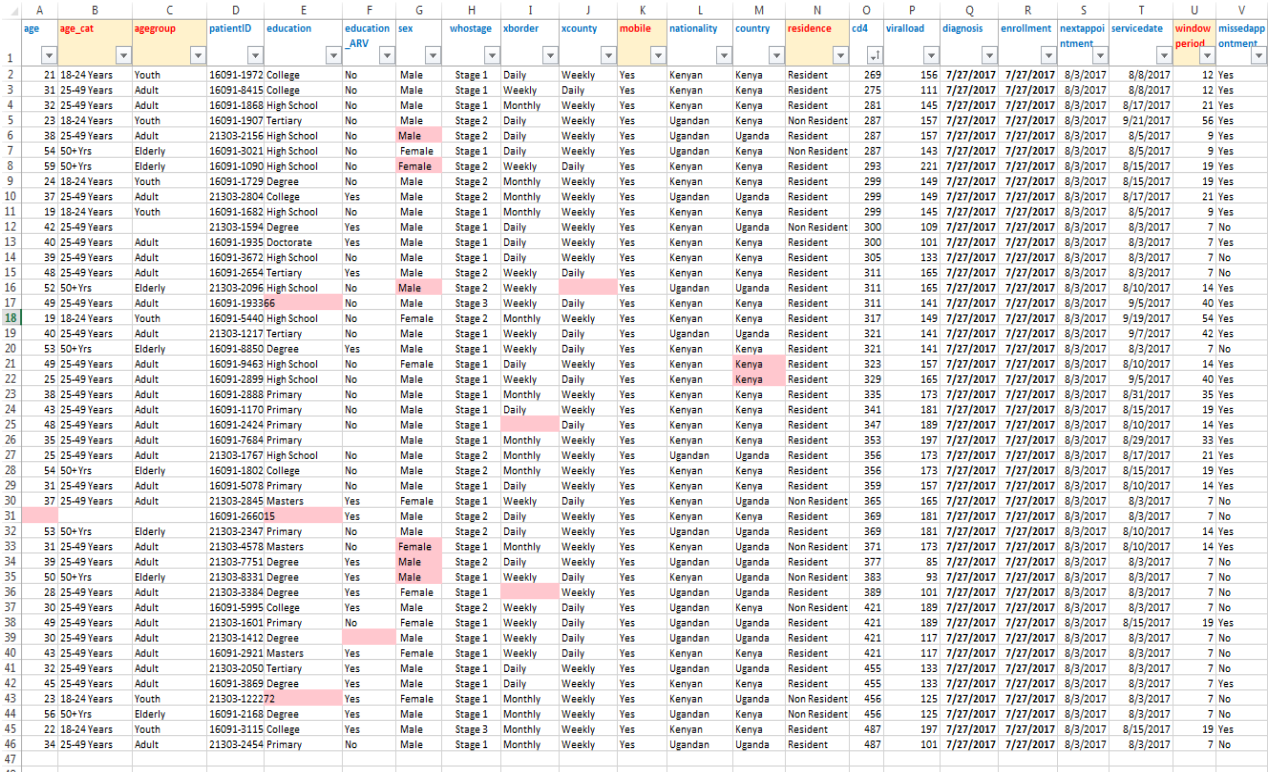
- **Step 5: Fix Problems with the Data -** All data is dirty, has problems. What is or isn't a problem varies with the data mining technique. For some, such as decision trees, missing values, and outliers do not cause too much trouble. In removing redundant data, the Kenya and Uganda data sets combined into one file by mapping similar variables, inconsistent values that were erroneously spelled were harmonized consequently, outliers and extreme values were removed, filling in missing values

FIGURE 27
Fixing problems with the data

	age	patientID	education	education_ARV	sex	whostage	xborder	xcounty	country	cd4	viralload	diagnosis	enrollment	nextappointment	servicedate	missedappointment	ltfu
2	21	16091-1972	College	No	Male	Stage 1	Daily	Weekly	Kenya	269	156	7/27/2017	7/27/2017	8/3/2017	8/8/2017	Yes	Yes
3	31	16091-8415	College	No	Male	Stage 1	Weekly	Daily	Kenya	275	111	7/27/2017	7/27/2017	8/3/2017	8/8/2017	Yes	Yes
4	32	16091-1868	High School	No	Male	Stage 1	Monthly	Weekly	Kenya	281	145	7/27/2017	7/27/2017	8/3/2017	8/17/2017	Yes	Yes
5	23	16091-1907	Tertiary	No	Male	Stage 2	Daily	Weekly	Kenya	287	157	7/27/2017	7/27/2017	8/3/2017	9/21/2017	Yes	No
6	38	21303-2156	High School	No	M	Stage 2	Daily	Weekly	Uganda	287	157	7/27/2017	7/27/2017	8/3/2017	8/5/2017	Yes	Yes
7	54	16091-3021	High School	No	Female	Stage 1	Daily	Weekly	Kenya	287	143	7/27/2017	7/27/2017	8/3/2017	8/5/2017	Yes	Yes
8	59	16091-1090	High School	No	F	Stage 2	Weekly	Daily	Kenya	293	221	7/27/2017	7/27/2017	8/3/2017	8/15/2017	Yes	Yes
9	24	16091-1729	Degree	No	Male	Stage 2	Monthly	Weekly	Kenya	299	149	7/27/2017	7/27/2017	8/3/2017	8/15/2017	Yes	Yes
10	37	21303-2804	College	Yes	Male	Stage 2	Monthly	Weekly	Uganda	299	149	7/27/2017	7/27/2017	8/3/2017	8/17/2017	Yes	Yes
11	19	16091-1682	High School	No	Male	Stage 1	Monthly	Weekly	Kenya	299	145	7/27/2017	7/27/2017	8/3/2017	8/5/2017	Yes	Yes
12	42	21303-1594	Degree	Yes	Male	Stage 1	Daily	Weekly	Uganda	300	109	7/27/2017	7/27/2017	8/3/2017	8/3/2017	No	No
13	40	16091-1935	Doctorate	Yes	Male	Stage 1	Daily	Weekly	Kenya	300	101	7/27/2017	7/27/2017	8/3/2017	8/3/2017	Yes	Yes
14	39	16091-3672	High School	No	Male	Stage 1	Daily	Weekly	Kenya	305	133	7/27/2017	7/27/2017	8/3/2017	8/3/2017	No	No
15	48	16091-2654	Tertiary	Yes	Male	Stage 2	Weekly	Daily	Kenya	311	165	7/27/2017	7/27/2017	8/3/2017	8/3/2017	No	No
16	52	21303-2096	High School	No	M	Stage 2	Weekly		Uganda	311	165	7/27/2017	7/27/2017	8/3/2017	8/10/2017	Yes	Yes
17	49	16091-193366		No	Male	Stage 3	Weekly	Daily	Kenya	311	141	7/27/2017	7/27/2017	8/3/2017	9/5/2017	Yes	Yes
18	19	16091-5440	High School	No	Female	Stage 2	Monthly	Weekly	Kenya	317	149	7/27/2017	7/27/2017	8/3/2017	9/19/2017	Yes	Yes
19	40	21303-1217	Tertiary	No	Male	Stage 1	Weekly	Daily	Uganda	321	141	7/27/2017	7/27/2017	8/3/2017	9/7/2017	Yes	Yes
20	53	16091-8850	Degree	Yes	Male	Stage 1	Weekly	Daily	Kenya	321	141	7/27/2017	7/27/2017	8/3/2017	8/3/2017	No	No
21	49	16091-9463	High School	No	Female	Stage 1	Daily	Weekly		323	157	7/27/2017	7/27/2017	8/3/2017	8/10/2017	Yes	Yes
22	25	16091-2899	High School	No	Male	Stage 1	Weekly	Daily		329	165	7/27/2017	7/27/2017	8/3/2017	9/5/2017	Yes	Yes
23	38	16091-2888	Primary	No	Male	Stage 1	Monthly	Weekly	Kenya	335	173	7/27/2017	7/27/2017	8/3/2017	8/31/2017	Yes	Yes
24	43	16091-1170	Primary	No	Male	Stage 1	Daily	Weekly	Kenya	341	181	7/27/2017	7/27/2017	8/3/2017	8/15/2017	Yes	Yes
25	48	16091-2424	Primary	No	Male	Stage 1		Daily	Kenya	347	189	7/27/2017	7/27/2017	8/3/2017	8/10/2017	Yes	Yes
26	35	16091-7684	Primary		Male	Stage 1	Monthly	Weekly	Kenya	353	197	7/27/2017	7/27/2017	8/3/2017	8/29/2017	Yes	No
27	25	21303-1767	High School	No	Male	Stage 2	Monthly	Weekly	Uganda	356	173	7/27/2017	7/27/2017	8/3/2017	8/17/2017	Yes	Yes
28	54	16091-1802	College	No	Male	Stage 2	Monthly	Weekly	Kenya	356	173	7/27/2017	7/27/2017	8/3/2017	8/15/2017	Yes	Yes
29	31	16091-5078	Primary	No	Male	Stage 1	Daily	Weekly	Kenya	359	157	7/27/2017	7/27/2017	8/3/2017	8/10/2017	Yes	Yes
30	37	21303-2845	Masters	Yes	Female	Stage 1	Weekly	Daily	Uganda	365	165	7/27/2017	7/27/2017	8/3/2017	8/3/2017	No	No
31		16091-266015		Yes	Male	Stage 2	Daily	Weekly	Kenya	369	181	7/27/2017	7/27/2017	8/3/2017	8/3/2017	No	No
32	53	21303-2347	Primary	No	Male	Stage 2	Daily	Weekly	Uganda	369	181	7/27/2017	7/27/2017	8/3/2017	8/10/2017	Yes	Yes
33	31	21303-4578	Masters	No	F	Stage 1	Monthly	Weekly	Uganda	371	173	7/27/2017	7/27/2017	8/3/2017	8/10/2017	Yes	Yes
34	39	21303-7751	Degree	Yes	M	Stage 2	Daily	Weekly	Uganda	377	85	7/27/2017	7/27/2017	8/3/2017	8/3/2017	No	No
35	50	21303-8331	Degree	Yes	M	Stage 1	Weekly	Daily	Uganda	383	93	7/27/2017	7/27/2017	8/3/2017	8/3/2017	No	No

- **Step 6: Transform Data to Bring Information to the Surface -** adding derived fields to bring information to the surface. It may also involve removing outliers, binning numeric variables, grouping classes for categorical variables, applying transformations. In Generalization and Discretization, the age variable was grouped and categorized into youth, adult and elderly, the Mobility was derived as a person crossing the border or county, the Residence of the clients was determined by nationality and country of enrolment and the Window period was calculated from date of initial ARV service and enrollment date

FIGURE 28
Data Transformation



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	age	age_cat	agegroup	patientID	education	education_ARV	sex	whostage	xborder	xcounty	mobile	nationality	country	residence	cd4	viralload	diagnosis	enrollment	nextappointment	servicedate	window period	missedappointment
2	21	18-24 Years	Youth	16091-1972	College	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	269	156	7/27/2017	7/27/2017	8/3/2017	8/8/2017	12	Yes
3	31	25-49 Years	Adult	16091-8415	College	No	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Kenya	Resident	275	111	7/27/2017	7/27/2017	8/3/2017	8/8/2017	12	Yes
4	32	25-49 Years	Adult	16091-1868	High School	No	Male	Stage 1	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	281	145	7/27/2017	7/27/2017	8/3/2017	8/17/2017	21	Yes
5	23	18-24 Years	Youth	16091-1907	Tertiary	No	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Kenya	Non Resident	287	157	7/27/2017	7/27/2017	8/3/2017	9/21/2017	56	Yes
6	38	25-49 Years	Adult	21303-2156	High School	No	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Uganda	Resident	287	157	7/27/2017	7/27/2017	8/3/2017	8/5/2017	9	Yes
7	54	50+Yrs	Elderly	16091-3021	High School	No	Female	Stage 1	Daily	Weekly	Yes	Ugandan	Kenya	Non Resident	287	143	7/27/2017	7/27/2017	8/3/2017	8/5/2017	9	Yes
8	59	50+Yrs	Elderly	16091-1090	High School	No	Female	Stage 2	Weekly	Daily	Yes	Kenyan	Kenya	Resident	293	221	7/27/2017	7/27/2017	8/3/2017	8/15/2017	19	Yes
9	24	18-24 Years	Youth	16091-1729	Degree	No	Male	Stage 2	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	299	149	7/27/2017	7/27/2017	8/3/2017	8/15/2017	19	Yes
10	37	25-49 Years	Adult	21303-2804	College	Yes	Male	Stage 2	Monthly	Weekly	Yes	Ugandan	Uganda	Resident	299	149	7/27/2017	7/27/2017	8/3/2017	8/17/2017	21	Yes
11	19	18-24 Years	Youth	16091-1682	High School	No	Male	Stage 1	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	299	145	7/27/2017	7/27/2017	8/3/2017	8/5/2017	9	Yes
12	42	25-49 Years		21303-1594	Degree	Yes	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Uganda	Non Resident	300	109	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
13	40	25-49 Years	Adult	16091-1935	Doctorate	Yes	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	300	101	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	Yes
14	39	25-49 Years	Adult	16091-3672	High School	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	305	133	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
15	48	25-49 Years	Adult	16091-2654	Tertiary	Yes	Male	Stage 2	Weekly	Daily	Yes	Kenyan	Kenya	Resident	311	165	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
16	52	50+Yrs	Elderly	21303-2096	High School	No	Male	Stage 2	Weekly		Yes	Ugandan	Uganda	Resident	311	165	7/27/2017	7/27/2017	8/3/2017	8/10/2017	14	Yes
17	49	25-49 Years	Adult	16091-193366		No	Male	Stage 3	Weekly	Daily	Yes	Kenyan	Kenya	Resident	311	141	7/27/2017	7/27/2017	8/3/2017	9/5/2017	40	Yes
18	19	18-24 Years	Youth	16091-5440	High School	No	Female	Stage 2	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	317	149	7/27/2017	7/27/2017	8/3/2017	9/19/2017	54	Yes
19	40	25-49 Years	Adult	21303-1217	Tertiary	No	Male	Stage 1	Weekly	Daily	Yes	Ugandan	Uganda	Resident	321	141	7/27/2017	7/27/2017	8/3/2017	9/7/2017	42	Yes
20	53	50+Yrs	Elderly	16091-8850	Degree	Yes	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Kenya	Resident	321	141	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
21	49	25-49 Years	Adult	16091-9463	High School	No	Female	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	323	157	7/27/2017	7/27/2017	8/3/2017	8/10/2017	14	Yes
22	25	25-49 Years	Adult	16091-2899	High School	No	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Kenya	Resident	329	165	7/27/2017	7/27/2017	8/3/2017	9/5/2017	40	Yes
23	38	25-49 Years	Adult	16091-2888	Primary	No	Male	Stage 1	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	335	173	7/27/2017	7/27/2017	8/3/2017	8/31/2017	35	Yes
24	43	25-49 Years	Adult	16091-1170	Primary	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	341	181	7/27/2017	7/27/2017	8/3/2017	8/15/2017	19	Yes
25	48	25-49 Years	Adult	16091-2424	Primary	No	Male	Stage 1		Daily	Yes	Kenyan	Kenya	Resident	347	189	7/27/2017	7/27/2017	8/3/2017	8/10/2017	14	Yes
26	35	25-49 Years	Adult	16091-7684	Primary		Male	Stage 1	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	353	197	7/27/2017	7/27/2017	8/3/2017	8/29/2017	33	Yes
27	25	25-49 Years	Adult	21303-1767	High School	No	Male	Stage 2	Monthly	Weekly	Yes	Ugandan	Uganda	Resident	356	173	7/27/2017	7/27/2017	8/3/2017	8/17/2017	21	Yes
28	54	50+Yrs	Elderly	16091-1802	College	No	Male	Stage 2	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	356	173	7/27/2017	7/27/2017	8/3/2017	8/15/2017	19	Yes
29	31	25-49 Years	Adult	16091-5078	Primary	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	359	157	7/27/2017	7/27/2017	8/3/2017	8/10/2017	14	Yes
30	37	25-49 Years	Adult	21303-2845	Masters	Yes	Female	Stage 1	Weekly	Daily	Yes	Kenyan	Uganda	Non Resident	365	165	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
31				16091-266015		Yes	Male	Stage 2	Daily	Weekly	Yes	Kenyan	Kenya	Resident	369	181	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
32	53	50+Yrs	Elderly	21303-2347	Primary	No	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Uganda	Resident	369	181	7/27/2017	7/27/2017	8/3/2017	8/10/2017	14	Yes
33	31	25-49 Years	Adult	21303-4578	Masters	No	Female	Stage 1	Monthly	Weekly	Yes	Kenyan	Uganda	Non Resident	371	173	7/27/2017	7/27/2017	8/3/2017	8/10/2017	14	Yes
34	39	25-49 Years	Adult	21303-7751	Degree	Yes	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Uganda	Resident	377	85	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
35	50	50+Yrs	Elderly	21303-8331	Degree	Yes	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Uganda	Non Resident	383	93	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
36	28	25-49 Years	Adult	21303-3384	Degree	Yes	Female	Stage 1		Weekly	Yes	Ugandan	Uganda	Resident	389	101	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
37	30	25-49 Years	Adult	16091-5995	College	Yes	Male	Stage 2	Weekly	Daily	Yes	Ugandan	Kenya	Non Resident	421	189	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
38	49	25-49 Years	Adult	21303-1601	Primary	No	Female	Stage 1	Weekly	Daily	Yes	Ugandan	Uganda	Resident	421	189	7/27/2017	7/27/2017	8/3/2017	8/15/2017	19	Yes
39	30	25-49 Years	Adult	21303-1412	Degree		Male	Stage 1	Weekly	Daily	Yes	Ugandan	Uganda	Resident	421	117	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
40	43	25-49 Years	Adult	16091-2921	Masters	Yes	Female	Stage 1	Weekly	Daily	Yes	Kenyan	Kenya	Resident	421	117	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
41	32	25-49 Years	Adult	21303-2050	Tertiary	Yes	Male	Stage 1	Daily	Weekly	Yes	Ugandan	Uganda	Resident	455	133	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
42	45	25-49 Years	Adult	16091-3869	Degree	Yes	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	455	133	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	Yes
43	23	18-24 Years	Youth	21303-122272		Yes	Female	Stage 1	Monthly	Weekly	Yes	Kenyan	Uganda	Non Resident	456	125	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
44	56	50+Yrs	Elderly	16091-2168	Degree	Yes	Male	Stage 1	Monthly	Weekly	Yes	Ugandan	Kenya	Non Resident	456	125	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
45	22	18-24 Years	Youth	16091-3115	College	Yes	Male	Stage 3	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	487	197	7/27/2017	7/27/2017	8/3/2017	8/15/2017	19	Yes
46	34	25-49 Years	Adult	21303-2454	Primary	No	Male	Stage 1	Monthly	Weekly	Yes	Ugandan	Uganda	Resident	487	101	7/27/2017	7/27/2017	8/3/2017	8/3/2017	7	No
47																						

- **Step 7: Building the Model -** Data Analytics Tools that intersect multiple reviews are Tanagra, Orange, KNIME, Weka, and Rapid Miner. The New Stack website discusses six (6) data mining toolkits: Orange, Weka, Rapid Miner, JHepWork, and KNIME (Goopta, 2014) whereas Tech Source website discusses five (5) data mining toolkits: Rapid Miner, Weka, Orange, R, KNIME, and NTLK (Auza, 2010). Those in academic or practitioner sources were; Orange, Tanagra, Rapid Miner, Weka, and KNIME.

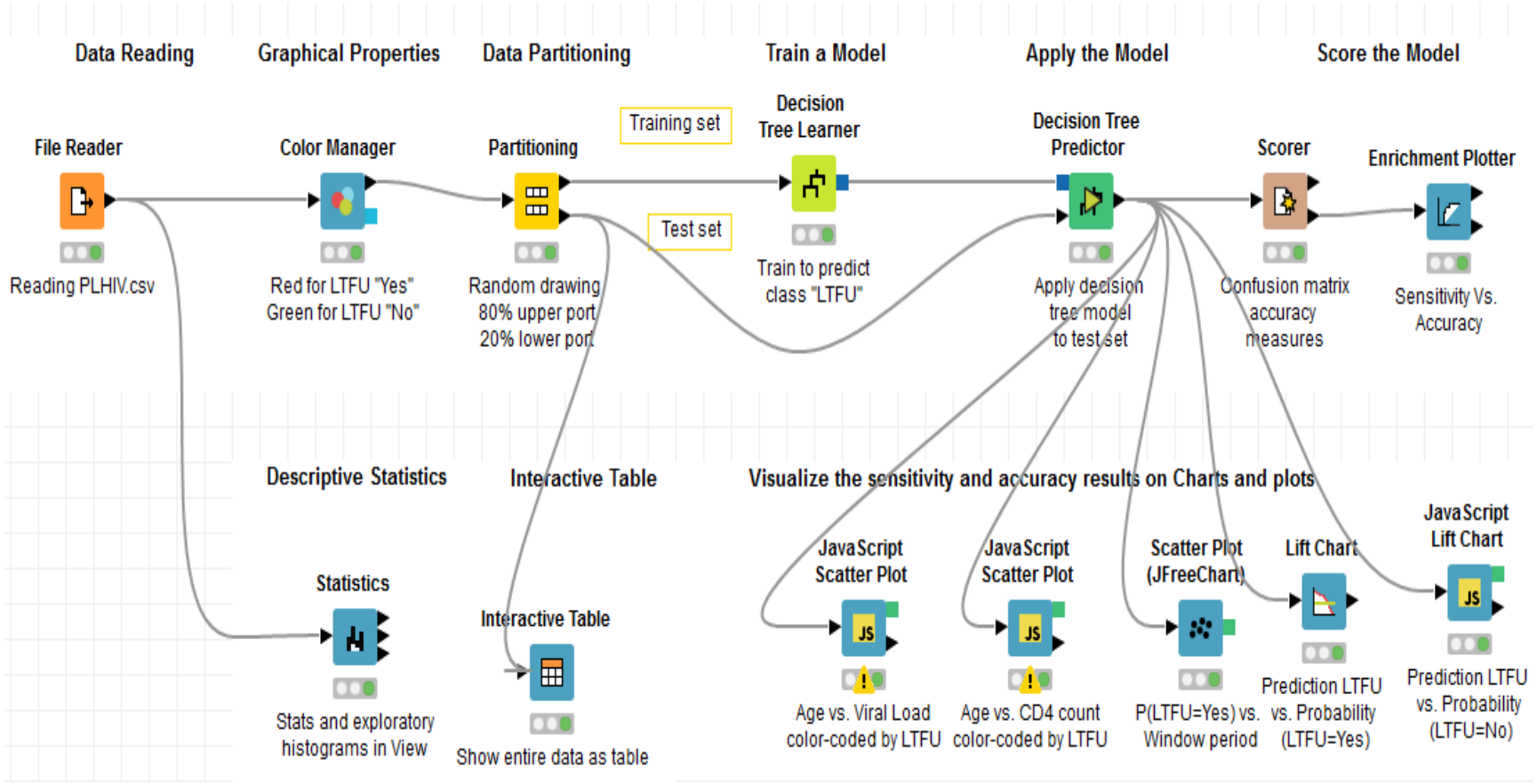
TABLE 5
Data Mining and data Analytics tool selection comparison matrix

Methods	Orange	Tenagra	Rapid Miner	R	WEKA	KNIME
K-means Clustering	Yes	Yes	Yes	Yes	Yes	Yes
Association rule mining	Yes	Yes	Yes	Yes	Yes	Yes
Linear regression	Yes	Yes	Yes	Yes	Yes	Yes
Logistic regression	Yes	Yes	Yes	Yes	Yes	Yes
Naïve Bayesian classifiers	Yes	Yes	Yes	Yes	Yes	Yes
Decision Tree	Yes	Yes	Yes	Yes	Yes	Yes
Time Series Analysis	No	No	Some	Yes	Yes	Yes
Text Analytics	Yes	No	Yes	Yes	Yes	Yes
Big Data Processing	No	No	No	Yes	Yes	Yes
Visual Workflows	Yes	Yes	Yes	No	Yes	Yes
Integrated GUI	Yes	Yes	Yes	No	No	Yes
Ease of Use	Yes	Yes	No	Yes	No	Yes

Source: Wimmer, 2015: A Comparison of Open Source Tools for Data Science

KNIME integrates functionality of open source projects covering all major areas of Data Analytics. Plug-ins allow creation of wrappers for former data analysis tools without having to modify executables.

FIGURE 29
LTFU Learner and Predictor designed and developed



The overall goal is to develop a prediction model based on parameters (demographics, profiling and service data) resulting from literature review to predict what could be the possible outcome; is a PLHIV enrolled on treatment a possible LTFU? What immediate measures should be taken to ensure adherence and retention to ARV drugs?

1. **The chosen dataset** was data already available at two participating facilities - not used for mining but the learner node used to train the model which included three files, master patients enrollment list, facility profile file and ART/CCC service file (Daily Activity Register). Loading the data was through the file reader node.
2. **Filtered variables** (removing unwanted columns using the "column filter function" that excludes unwanted and includes the required columns) linked with the client's unique identifier to produce a joined table from entries made on the LTFU system prototype.
3. **Stratified random sampling** was used on the partitioning node to partition the datasets where the whole group was divided into different datasets based on categories (mobility) called strata with a relative 80% as the training set and the remaining 20% for the prediction.
4. **Once all the nodes of the complete workflow** were configured, the researcher connected the partitioned dataset with learner and directed the first dataset to build the model and then passed the model and second dataset to the predictor to predict the LTFU class.
5. Under **Prediction,** the Decision Tree Learner - first partition of the data went to the decision tree learner to build the model and learn it through the data. The model was sent to the Decision Tree Predictor with the second partition data to predict the LTFU class. Data mining algorithms like decision trees J48, Naïve Bayes and PART decisions were used to mine the data and generate predictions of possible loss-to-follow-up and a set of evaluation results analyses and validated based on documented good practices.

6. The researcher added the **scorer** at the end of the workflow to gather statistics for evaluating the model (know how precise or accurate the model is, how good it is in predicting the values). It gives the confusion matrix (correctly and wrongly classified classes), accuracy of the model, error rate of the model and Cohen's Kappa (K) statistic.
7. **Metrics** such as False Positive, False Negative, precision and recall, F-Measure would measure how precise/accurate the prediction model is. Instances with low accuracy, precision would require refinement of the model to a better one through noise elimination on the training dataset. Comparison of the Kohen's Kappa Statistic, Mean absolute error, Root mean square error, Relative absolute error, F-Measure and the ROC Area from the data mining algorithms gave an indication of which algorithm would be stronger in predicting the LTFU class.
8. **Data visualization** is useful to explore the data visually or create graphics for instance those used in reports to describe the data and the underlying system. It is always an important step preceding the analytics phase as it is useful to visualize the data to check for obvious data clusters, interesting outliers, or statistical properties before proceeding with the application of analytics techniques. In this study it was based on parameters (descriptive statistics, interactive tables, scatter plots, decision tree, Lift, gain and the ROC chart) demonstrating how PLHIVs are classified/distributed.

FIGURE 30
Color Manager Node demonstration of LTFU categorization

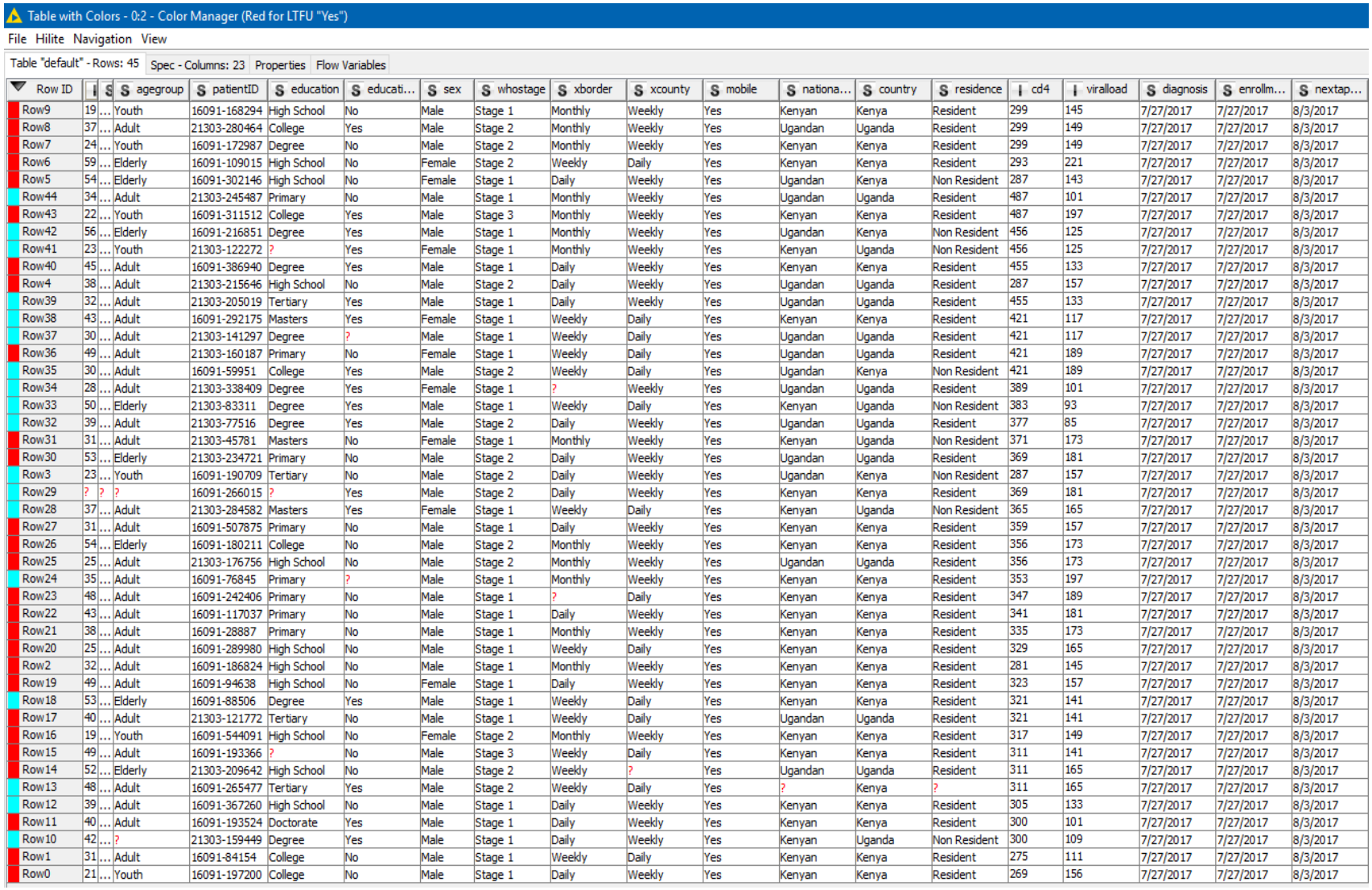


Table with Colors - 0:2 - Color Manager (Red for LTFU "Yes")

File Hilite Navigation View

Table "default" - Rows: 45 | Spec - Columns: 23 | Properties | Flow Variables

Row ID	I	S	agegroup	patientID	education	educati...	sex	whostage	xborder	xcounty	mobile	nationa...	country	residence	cd4	viralload	diagnosis	enrollm...	nextap...
Row9	19	...	Youth	16091-168294	High School	No	Male	Stage 1	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	299	145	7/27/2017	7/27/2017	8/3/2017
Row8	37	...	Adult	21303-280464	College	Yes	Male	Stage 2	Monthly	Weekly	Yes	Ugandan	Uganda	Resident	299	149	7/27/2017	7/27/2017	8/3/2017
Row7	24	...	Youth	16091-172987	Degree	No	Male	Stage 2	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	299	149	7/27/2017	7/27/2017	8/3/2017
Row6	59	...	Elderly	16091-109015	High School	No	Female	Stage 2	Weekly	Daily	Yes	Kenyan	Kenya	Resident	293	221	7/27/2017	7/27/2017	8/3/2017
Row5	54	...	Elderly	16091-302146	High School	No	Female	Stage 1	Daily	Weekly	Yes	Ugandan	Kenya	Non Resident	287	143	7/27/2017	7/27/2017	8/3/2017
Row44	34	...	Adult	21303-245487	Primary	No	Male	Stage 1	Monthly	Weekly	Yes	Ugandan	Uganda	Resident	487	101	7/27/2017	7/27/2017	8/3/2017
Row43	22	...	Youth	16091-311512	College	Yes	Male	Stage 3	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	487	197	7/27/2017	7/27/2017	8/3/2017
Row42	56	...	Elderly	16091-216851	Degree	Yes	Male	Stage 1	Monthly	Weekly	Yes	Ugandan	Kenya	Non Resident	456	125	7/27/2017	7/27/2017	8/3/2017
Row41	23	...	Youth	21303-122272	?	Yes	Female	Stage 1	Monthly	Weekly	Yes	Kenyan	Uganda	Non Resident	456	125	7/27/2017	7/27/2017	8/3/2017
Row40	45	...	Adult	16091-386940	Degree	Yes	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	455	133	7/27/2017	7/27/2017	8/3/2017
Row4	38	...	Adult	21303-215646	High School	No	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Uganda	Resident	287	157	7/27/2017	7/27/2017	8/3/2017
Row39	32	...	Adult	21303-205019	Tertiary	Yes	Male	Stage 1	Daily	Weekly	Yes	Ugandan	Uganda	Resident	455	133	7/27/2017	7/27/2017	8/3/2017
Row38	43	...	Adult	16091-292175	Masters	Yes	Female	Stage 1	Weekly	Daily	Yes	Kenyan	Kenya	Resident	421	117	7/27/2017	7/27/2017	8/3/2017
Row37	30	...	Adult	21303-141297	Degree	?	Male	Stage 1	Weekly	Daily	Yes	Ugandan	Uganda	Resident	421	117	7/27/2017	7/27/2017	8/3/2017
Row36	49	...	Adult	21303-160187	Primary	No	Female	Stage 1	Weekly	Daily	Yes	Ugandan	Uganda	Resident	421	189	7/27/2017	7/27/2017	8/3/2017
Row35	30	...	Adult	16091-59951	College	Yes	Male	Stage 2	Weekly	Daily	Yes	Ugandan	Kenya	Non Resident	421	189	7/27/2017	7/27/2017	8/3/2017
Row34	28	...	Adult	21303-338409	Degree	Yes	Female	Stage 1	?	Weekly	Yes	Ugandan	Uganda	Resident	389	101	7/27/2017	7/27/2017	8/3/2017
Row33	50	...	Elderly	21303-83311	Degree	Yes	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Uganda	Non Resident	383	93	7/27/2017	7/27/2017	8/3/2017
Row32	39	...	Adult	21303-77516	Degree	Yes	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Uganda	Resident	377	85	7/27/2017	7/27/2017	8/3/2017
Row31	31	...	Adult	21303-45781	Masters	No	Female	Stage 1	Monthly	Weekly	Yes	Kenyan	Uganda	Non Resident	371	173	7/27/2017	7/27/2017	8/3/2017
Row30	53	...	Elderly	21303-234721	Primary	No	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Uganda	Resident	369	181	7/27/2017	7/27/2017	8/3/2017
Row3	23	...	Youth	16091-190709	Tertiary	No	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Kenya	Non Resident	287	157	7/27/2017	7/27/2017	8/3/2017
Row29	?	?	?	16091-266015	?	Yes	Male	Stage 2	Daily	Weekly	Yes	Kenyan	Kenya	Resident	369	181	7/27/2017	7/27/2017	8/3/2017
Row28	37	...	Adult	21303-284582	Masters	Yes	Female	Stage 1	Weekly	Daily	Yes	Kenyan	Uganda	Non Resident	365	165	7/27/2017	7/27/2017	8/3/2017
Row27	31	...	Adult	16091-507875	Primary	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	359	157	7/27/2017	7/27/2017	8/3/2017
Row26	54	...	Elderly	16091-180211	College	No	Male	Stage 2	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	356	173	7/27/2017	7/27/2017	8/3/2017
Row25	25	...	Adult	21303-176756	High School	No	Male	Stage 2	Monthly	Weekly	Yes	Ugandan	Uganda	Resident	356	173	7/27/2017	7/27/2017	8/3/2017
Row24	35	...	Adult	16091-76845	Primary	?	Male	Stage 1	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	353	197	7/27/2017	7/27/2017	8/3/2017
Row23	48	...	Adult	16091-242406	Primary	No	Male	Stage 1	?	Daily	Yes	Kenyan	Kenya	Resident	347	189	7/27/2017	7/27/2017	8/3/2017
Row22	43	...	Adult	16091-117037	Primary	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	341	181	7/27/2017	7/27/2017	8/3/2017
Row21	38	...	Adult	16091-28887	Primary	No	Male	Stage 1	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	335	173	7/27/2017	7/27/2017	8/3/2017
Row20	25	...	Adult	16091-289980	High School	No	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Kenya	Resident	329	165	7/27/2017	7/27/2017	8/3/2017
Row2	32	...	Adult	16091-186824	High School	No	Male	Stage 1	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	281	145	7/27/2017	7/27/2017	8/3/2017
Row19	49	...	Adult	16091-94638	High School	No	Female	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	323	157	7/27/2017	7/27/2017	8/3/2017
Row18	53	...	Elderly	16091-88506	Degree	Yes	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Kenya	Resident	321	141	7/27/2017	7/27/2017	8/3/2017
Row17	40	...	Adult	21303-121772	Tertiary	No	Male	Stage 1	Weekly	Daily	Yes	Ugandan	Uganda	Resident	321	141	7/27/2017	7/27/2017	8/3/2017
Row16	19	...	Youth	16091-544091	High School	No	Female	Stage 2	Monthly	Weekly	Yes	Kenyan	Kenya	Resident	317	149	7/27/2017	7/27/2017	8/3/2017
Row15	49	...	Adult	16091-193366	?	No	Male	Stage 3	Weekly	Daily	Yes	Kenyan	Kenya	Resident	311	141	7/27/2017	7/27/2017	8/3/2017
Row14	52	...	Elderly	21303-209642	High School	No	Male	Stage 2	Weekly	?	Yes	Ugandan	Uganda	Resident	311	165	7/27/2017	7/27/2017	8/3/2017
Row13	48	...	Adult	16091-265477	Tertiary	Yes	Male	Stage 2	Weekly	Daily	Yes	?	Kenya	?	311	165	7/27/2017	7/27/2017	8/3/2017
Row12	39	...	Adult	16091-367260	High School	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	305	133	7/27/2017	7/27/2017	8/3/2017
Row11	40	...	Adult	16091-193524	Doctorate	Yes	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	300	101	7/27/2017	7/27/2017	8/3/2017
Row10	42	...	?	21303-159449	Degree	Yes	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Uganda	Non Resident	300	109	7/27/2017	7/27/2017	8/3/2017
Row1	31	...	Adult	16091-84154	College	No	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Kenya	Resident	275	111	7/27/2017	7/27/2017	8/3/2017
Row0	21	...	Youth	16091-197200	College	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Kenya	Resident	269	156	7/27/2017	7/27/2017	8/3/2017

CHAPTER FOUR
RESEARCH FINDINGS AND DISCUSSIONS

4.1 Introduction

This chapter presents the findings of the study. The purpose of this study was to understand the factors that motivate mobile PLHIVs to default anti-retroviral drugs (ARV), understand the causes or effects of these factors to overcome the difficulty with increasing rates of loss-to-follow-up at two participating health facilities at the cross-border study site. The study seeks to discover what is required in designing a tool that utilizes data mining techniques to mine existing PLHIV daily activity register data to predict if a client would be a possible loss-to-follow-up. the researcher also seeks to determine what is required in developing a data mining and analytics tool to predict possible loss-to-follow-up among mobile PLHIVs which was then be validated and implemented at the facilities and thus provide a new interpretation of tracing mobile PLHIVs displaying transnational service access characteristics.

4.2 Results and findings

4.2.1 Factors that motivate PLHIVs to default ARV drugs and the levels of loss-to-follow-up

The reviewed literature examined factors that motivate people living with HIV not to adhere to ART, the levels of adherence at the participating health facilities. It showed that from there are many predisposing characteristics, enabling resources and need factors that influence adherence to ART. The review identified; *window period before initiation to care and treatment, first adherence patterns, CD4+ cell counts, unsuppressed viral load, age at enrolment, gender, educational level and unofficial residents with smaller social network size* as the main variables affecting adherence to ART. The eight have more weight and informed the design and development of the tool for predicting possible loss-to-follow-up amongst people living with HIV.

TABLE 6
Selected Factors Motivating PLHIVs to Default ARV Drugs

Predisposing characteristics • Beliefs, lack of acceptance and knowledge, stigma and fears, age at initiation, gender, marital status, educational level, religion, occupation type	Reviewed literature and discussions with relevant stake holders shows that these are the main predictors for loss-to-follow-up among people living with HIV	**Selected factors** 1. Window period before initiation to care and treatment 2. First ARV adherence patterns (definition of missed appointments) 3. CD4+ cell counts (from PLHIV viral load tests) 4. Unsuppressed viral load (response to treatment, behavior and nutrition) 5. Gender (females enrolled adhere to ARV treatment more than men) 6. Age at enrollment into care (vulnerable populations are between 15 to 49 years) 7. Educational level (knowledge on drug availability, adherence and treatment) 8. Unofficial residents with small social network size (client mobility, migrants)
Enabling Resources • Social and financial support issues, health worker attitudes, access to ART, poor documentation of services, defaulter tracing, occupation status, transfer in status, tracing status, location of ART management, time on ART		
Need Factors • Initial weight at enrollment, WHO stage at ART initiation, baseline CD4 count, weight at follow up, ARV drug side effects, tuberculosis, viral load suppression, ART continuation decision		

There are existing and operationalized national reporting tools (Kenya and Uganda) that are currently being used in the collection of service access by PLHIVs enrolled in care and treatment. They are the daily activity registers referred to as the blue card in Uganda. Majority of

the variables collected are similar but for the small difference in the specific per country labels used to refer to them.

Reviewing the two tools together revealed that the following Predictor Variables; *Unique ART/CCC number, HIV diagnosis date, Enrollment Date, CD4 count, Viral load, Gender (M/F), Age in years, WHO Stage and Education on ARV drugs* are available in both DAR tools. These Existing MoH National (Kenya/Uganda) tools lack variables that capture referred and walk-in client's nationality, residence status, mobility (cross-border, cross-county/district in a country) and their target population (key or priority) details. Through joint cross-border meetings with the health teams from the participating health facilities through their county and district health management teams (C/DHMTs) together with representation from their technical working groups (TWGs), discussions revolved around the clinical officers, nurses and clinicians at the facilities already having a lot of documentation to complete at routine client visits, the registers they have to fill to report to the government and other bilateral organization has been on the increase and introduction of parallel tools to capture data for the purpose of this study would end up being ignored or if filled, low-quality information is provided, the tools did not have complete data that has integrity, verifiable if they are not monetarily motivated (strategies suggested include; supporting their snacks and refreshments during their weekly continuous medical education (CMEs) conducted every Wednesdays, providing them with airtime for communication and follow-up, reimbursing transport to those coming from far, purchasing computing and filing equipment for the participating health facilities).

FIGURE 31
Facility Client Profile Form

East Africa Cross-border Health System: **Profile Form** EACBHS tools 007

Country [1]: _____ Site [2]: _____ Facility: _____ Department: _____ MFL Code: _____ Reporting period: Month/Quarter: _____/ Year: _____

			3	4	5	8	9	10	11	12	13
Enrollment Date: (dd/mm/yyyy)	Client's name:	New/revisit	Registration number / Unique Client ID:	Nationality:	Contact telephone number:	Village/Main hotspot affiliation, contact/Support person (in case of follow-up) , land mark:	Is this (town/community) your usual place of residence?	If No, where is your usual place of residence ?	Target Population	On average how often would you say you move across the border in a month?	On average how often would you say you move to another county/District within the country in a month?
___/___/____											
___/___/____											

For the purpose of this study, the researcher introduced a "Facility Profile Form" to collect missing data at both facilities which would be would be linked to existing service register data using "*Unique Client ID/Registration number*", service date and the client name through a facility data linking and extraction form. Additional mobility qualification variables for instance "village or main hotspot of affiliation of the client" that identifies the target category of the clients key populations (FSWs, MSMs and PWIDs) operating at hotspots and dens whereas priority populations (LDTDs, CFAs, VWGs and Fishing populations) operating from villages. Here are instances where clients profiled are contacted at their areas of operation that would be different from their village of residence; the variable on "is this village/community your usual place of residence" would clearly demonstrate clients who move from their usual place of residence to other areas to work or for business. The variable on "if No, where is your usual place of residence" would provide details on where the client comes from.

Data collected during integrated health service outreaches (IHSO) conducted within the community and at hotspots when "those found to be HIV positive/reactive" are further analyzed by outreach venues; they would generate varying HIV positivity rates with the highest indicating areas with higher infection. A deeper understanding of those "tested for the first time" would indicate areas where clients have never been tested and do not know their HIV status. These would

be areas that people would infect each other without their knowledge. Positivity rates for the first time testers. In concluding, strategies such as improving early HIV diagnosis and the early initiation of ART, providing comprehensive nutrition support, implementing intensive monitoring using newly developed indicators, and tracking patients who do not return for their appointments are needed to ensure proper adherence to ARV drugs.

4.2.2 Designing a data mining/analytics tool to predict LTFUs among mobile PLHIVs.

Mweete Nglazi in his article "*Identification of losses to follow-up in a community-based antiretroviral therapy clinic in South Africa using a computerized pharmacy tracking system*" explains that computerized systems are used to track patients efficiently and are a potential solution for the rapid identification of patients potentially LTFU. These systems lack the capacity to mine the data stored in its repositories for knowledge generation. However, the utility of these systems depends on user friendliness, affordability, sustainability, stability, security and data ownership. He notes that sites across Africa employ a wide range of electronic information systems (*EHIS, EMRS*) to identify patients to be potentially traced (Nglazi *et al*. 2011).

National systems are a considerable effort by bilateral programs and the governments to track LTFUs, however, they require an uninterrupted electricity supply, a computer, and offsite back-up such a flash memory stick, cell phone, email or internet connection. In designing a system for predicting LTFU a number of factors should be considered; low cost, user friendliness, minimal staff training requirements and sustainability. The solution should be a simple predictor that can then inform the probable retention measures (CHWs/VHTs/locator information, support group leaders picking drugs on behalf of other PLHIVs, providing drugs enough for travel period, ability to replenish drugs at selected facilities across-the border) that could potentially be implemented in settings that are able to support the required infrastructure (Nglazi *et al*. 2011). Multiple factors

may be associated with LTFU and those determined through literature review in specific objective one was used in designing a tool for predicting possible LTFUs among PLHIVs at Busia border. This may enable more sophisticated data mining algorithms to identify patients who are potentially LTFU to be developed in due course. Outcomes may differ in clinics with different ART dispensing patterns.

4.2.3 Developing a data mining and analytics tool to predict possible LTFUs among mobile PLHIVs.

Selected variables evidenced by literature review were used to inform design of LTFUs System prototype with considerations of having it running on a portable device version (mobile ready). This was considered alongside best design practices from past studies in selecting the appropriate system development platform (examples CommCare, ODK Collect, Zoho Creator). The web-based software brings many advantages including big savings. Zoho Creator multiplies these benefits by providing the ability to create online database applications as needed, tailored-fit to specific needs instead of having a one-size, fit-all solution. Starting from scratch is just as easy with migration features, third-party integrations and ready-to-use solutions available in the app marketplace.

With Zoho, you can create a customized database application that meets your unique needs, add customized forms to your application, collect data from users of your application, analyze data that you have collected, perform certain tasks based on the data collected and prepare and publish useful reports that reflect the data you collect. It allows its users to design applications solely for use on your office computers, or you can make it accessible online and through mobile devices by anyone you choose. A well-researched and validated design process was adhered to in coming up with the prototype's user interface. Administration interfaces for Site, Country (*two at a radius of*

five kilometers form a site), County/District and participating facility. Mobile versions of the interfaces developed were generated and configured to communicate and synchronize with Zoho server which came in handy for the community health workers and the Village health teams working closely with the health facilities in enrolling clients to care and treatment.

4.2.4 Implementing and validating a data mining and analytics tool for predicting possible LTFUs among mobile PLHIVs.

In Step nine (9) of *Berry and Linnof's Data Mining Methodology* that involves deploying the data mining and analytics tool (enrollment of people living with HIV, LTFU prediction, and analytics flags) – CCC/ART clinic focal persons –the web application prototype running on Zoho platform servers was piloted at BCRH and BHCIV respectively with the D/CHMT teams (*Nelson Andanje – SCASCO/HIV Coordinator Busia County and Dr. Lam Mayende – DHE/IHS Busia District*) oriented on the purpose of implementing the tool at the selected participating facilities, the expected results and their impact in predicting possible LTFUs to enforce strategies that retain them on drugs longer. They supported the study and agreed that if well implemented and there is buy in by relevant MoH stakeholders, this tool simplified the work of linkage facilitators and retention teams.

Two PLHIV coordinators (*Alfayo Ogecha – PHO BCRH and Gertrude Nekesa – PHO BHC IV)* each from either side of the border were given access to the prototype. They were taken through its management and use and they entered data on it for a week (24^{th} to 30^{th} September 2017) and 15 (8 Kenyan and 7 Ugandan) PLHIV enrolment, profiling and service data were captured. Steps in preprocessing the data were followed (removing missing values, categorizing data into meaningful groups) and the data combined with the records in the training set (totaling to 60 records). The data was imported in to the data mining model, color coded with the color manager,

partitioned with the partition node, the model was trained through the decision tree learner node and possible LTFUs classified through the decision tree predictor.

FIGURE 18
Data captured for walk-in PLHIVs at Kenya and Uganda facilities

education	education_ARV	sex	whostage	xborder	xcounty	mobile	nationality	country	residence	cd4	viralload
Primary	No	Female	Stage 2	Yearly	Yearly	No	Ugandan	Uganda	Resident	421	181
Secondary	No	Female	Stage 1	Yearly	Yearly	No	Kenyan	Kenya	Resident	275	117
Tertiary	Yes	Female	Stage 3	Daily	Weekly	yes	Ugandan	Uganda	Non Resident	359	133
Tertiary	Yes	Female	Stage 2	Weekly	Yearly	No	Kenyan	Kenya	Resident	281	165
Tertiary	Yes	Female	Stage 2	Weekly	Weekly	yes	Ugandan	Uganda	Non Resident	353	189
Tertiary	Yes	Female	Stage 1	Weekly	Weekly	yes	Ugandan	Uganda	Non Resident	335	141
Tertiary	Yes	Female	Stage 2	Daily	Weekly	yes	Ugandan	Uganda	Resident	305	157
Primary	Yes	Female	Stage 1	Daily	Monthly	yes	Kenyan	Kenya	Resident	300	141
College	Yes	Female	Stage 1	Weekly	Weekly	yes	Kenyan	Kenya	Resident	341	143
Degree	Yes	Female	Stage 1	Daily	Weekly	yes	Ugandan	Uganda	Resident	421	173
Primary	No	Male	Stage 1	Yearly	Yearly	No	Kenyan	Kenya	Resident	455	125
College	No	Male	Stage 1	Weekly	Monthly	yes	Ugandan	Uganda	Non Resident	311	221
Degree	Yes	Female	Stage 1	Yearly	Yearly	No	Kenyan	Kenya	Resident	347	181
Primary	Yes	Male	Stage 1	Weekly	Monthly	yes	Kenyan	Kenya	Resident	311	125
College	Yes	Male	Stage 1	Daily	Weekly	yes	Kenyan	Kenya	Resident	323	133

FIGURE 19
User Interface Development

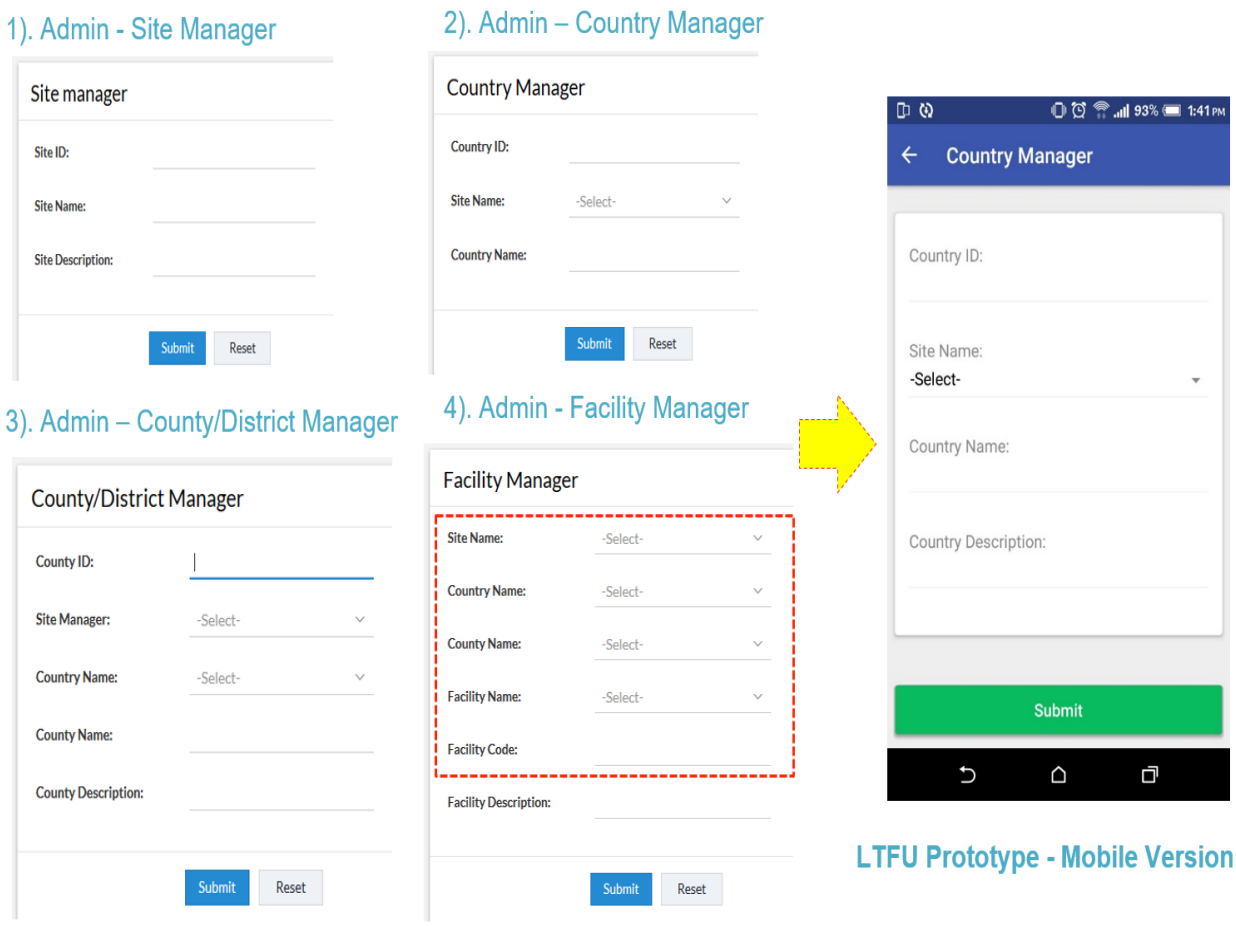
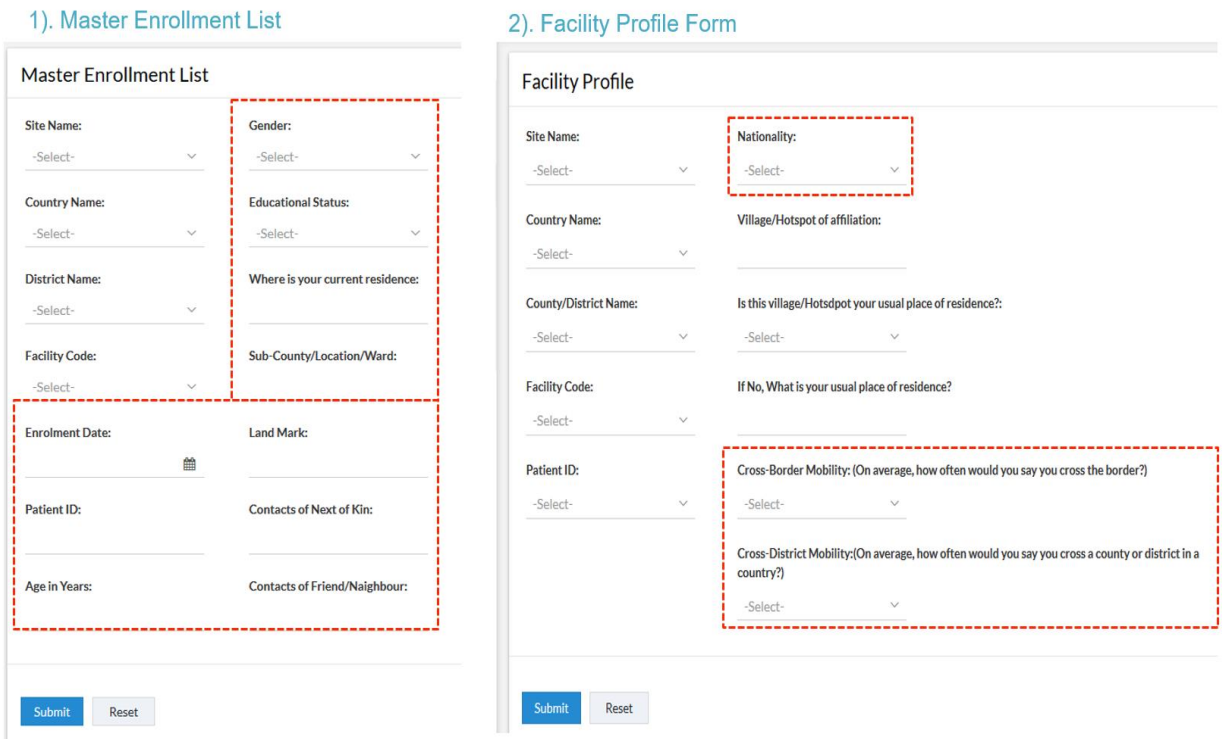
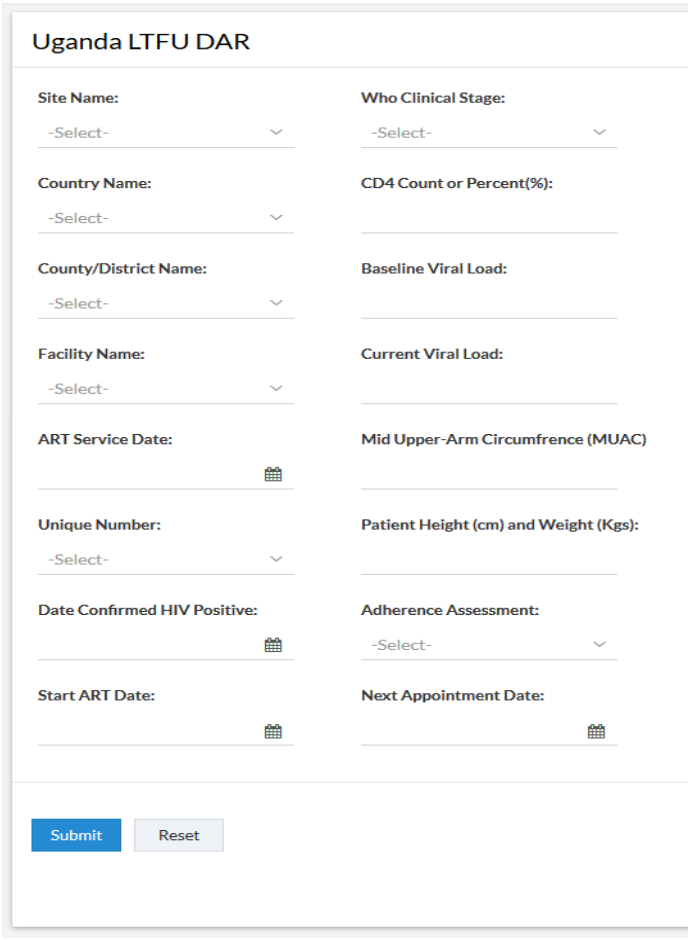


FIGURE 20
Master Facility List and Facility Profile form



- **Background information** variables are lookup selections from the system manager, enrolment and profiling forms
- **Patient unique number** used system wide to de-link the client from names which could compromise confidentiality
- Date confirmed HIV positive and ART start or initiation date used to determine **client's window period**
- **Adherence assessment** measures a clients while defaulting drugs giving good, fair,

FIGURE 21
Daily Activity Register



4.3 Discussion of Results

FIGURE 22
Classified data (Decision Tree Predictor - apply decision)

Classified Data - 0:4 - Decision Tree Predictor (Apply decision)

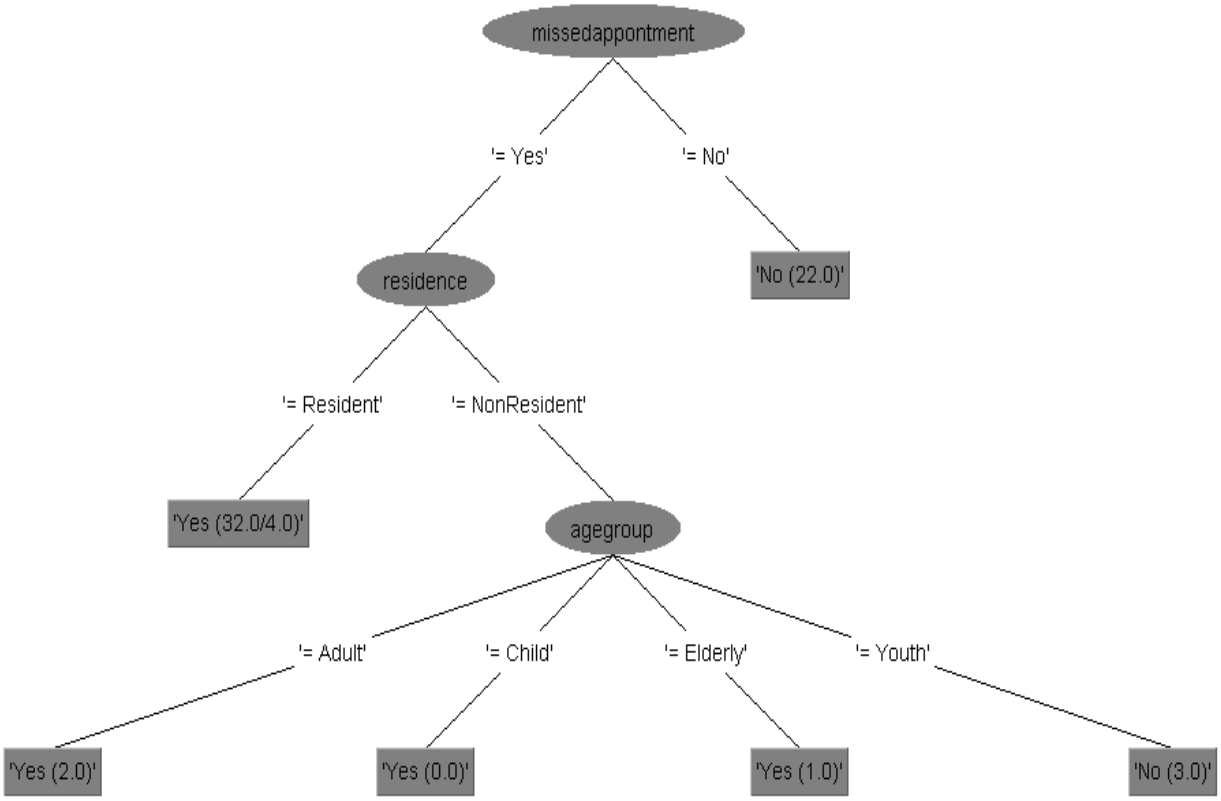
File Hilite Navigation View

Table "default" - Rows: 18 | Spec - Columns: 26 | Properties | Flow Variables

Row ID	patientID	agegroup	education_ARV	sex	whostage	xborder	xcounty	mobile	nationality	residence	missedappointment	ltfu	P (ltfu=Yes)	P (ltfu=No)	Prediction (ltfu)
Row8	16091-84154	Adult	No	Male	Stage 1	Weekly	Daily	Yes	Kenyan	Resident	Yes	Yes	0.786	0.214	Yes
Row7	16091-59951	Adult	Yes	Male	Stage 2	Weekly	Daily	Yes	Ugandan	Non Resident	No	No	0	1	No
Row59	16091-33431	Adult	Yes	Male	Stage 1	Daily	Weekly	yes	Kenyan	Resident	Yes	Yes	0.786	0.214	Yes
Row53	16091-24511	Adult	Yes	Female	Stage 1	Weekly	Weekly	yes	Kenyan	Resident	No	No	0	1	No
Row51	21303-67344	Youth	Yes	Female	Stage 2	Daily	Weekly	yes	Ugandan	Resident	No	No	0	1	No
Row49	21303-22454	Youth	Yes	Female	Stage 2	Weekly	Weekly	yes	Ugandan	Non Resident	Yes	No	0.786	0.214	Yes
Row47	21303-13009	Youth	Yes	Female	Stage 3	Daily	Weekly	yes	Ugandan	Non Resident	No	No	0	1	No
Row39	21303-121772	Adult	No	Male	Stage 1	Weekly	Daily	Yes	Ugandan	Resident	Yes	Yes	0.786	0.214	Yes
Row38	21303-77516	Adult	Yes	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Resident	No	No	0	1	No
Row37	21303-215646	Adult	No	Male	Stage 2	Daily	Weekly	Yes	Ugandan	Resident	Yes	Yes	0.786	0.214	Yes
Row3	16091-311512	Youth	Yes	Male	Stage 3	Monthly	Weekly	Yes	Kenyan	Resident	Yes	Yes	0.786	0.214	Yes
Row27	16091-266015	Adult	Yes	Male	Stage 2	Daily	Weekly	Yes	Kenyan	Resident	No	No	0	1	No
Row23	16091-302146	Elderly	No	Female	Stage 1	Daily	Weekly	Yes	Ugandan	Non Resident	Yes	Yes	0.786	0.214	Yes
Row19	16091-242406	Adult	No	Male	Stage 1	Daily	Daily	Yes	Kenyan	Resident	Yes	Yes	0.786	0.214	Yes
Row17	16091-386940	Adult	Yes	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Resident	Yes	Yes	0.786	0.214	Yes
Row16	16091-292175	Adult	Yes	Female	Stage 1	Weekly	Daily	Yes	Kenyan	Resident	No	No	0	1	No
Row13	16091-367260	Adult	No	Male	Stage 1	Daily	Weekly	Yes	Kenyan	Resident	No	No	0	1	No
Row1	16091-544091	Youth	No	Female	Stage 2	Monthly	Weekly	Yes	Kenyan	Resident	Yes	Yes	0.786	0.214	Yes

*The KNIME Prediction model classified the data based on the training set and predicted cases as possible loss-to-follow-up correctly and non-possible-loss-to-follow-up correctly. A single case of class **NO** was wrongly classified as of being of class **YES***

FIGURE 23
Decision Tree



Using the WEKA classifier plugin embodied on KNIME workspace, the researcher generated the above decision tree using J48 decision tree algorithm that demonstrates many interesting rules discussed below.

1. *If Missed appointment = **YES**, residence = **nonresident**, age group = **adult** OR Missed appointment = **YES**, residence = **nonresident**, age group = **child** OR Missed appointment = **YES**, residence = **nonresident**, age group = **elderly** THEN Possible Loss-to-follow-up*
2. *If Missed appointment = **YES**, residence status = **resident** THEN Possible Loss-to-follow-up*
3. *If Missed appointment = **YES**, residence status = **nonresident**, age group = **youth** THEN NOT Possible Loss-to-follow-up*
4. *If Missed appointment = **NO** THEN Not Possible Loss-to-follow-up*

FIGURE 24
PART Rule Decision List

```
=== Classifier model (full training set) ===

PART decision list
------------------

missedappontment = Yes AND
residence = Resident: Yes (32.0/4.0)

educationarv = Yes: No (22.0)

sex = Male: No (4.0/1.0)

: Yes (2.0)

Number of Rules  :      4
```

1. *Missed appointment = **YES** AND residence status = **resident THEN** possible Loss-To-Follow-up*
2. *Education ARV = **YES THEN** Not Possible Loss-To-Follow-up*

To evaluate the model the following measures were used to test their accuracy.

1. **Kappa Statistic:** This statistic is used to measure how strongly data items in the same class resemble each other. That is, the level of intra-class correlation. It is like the correlation coefficient.

TABLE 7
Kappa Statistic Measurements and Interpretation

Measurement (value or range)	**Interpretation**
0.0	Complete disagreement (do not resemble each other).
0.40 to 0.59	Moderate agreement
0.60 to 0.79	Substantial agreement
above 0.80	Outstanding agreement
1.0	Complete agreement (items strongly resemble each other)

2. **Mean Absolute Error (MAE):** Can be defined as sum of absolute errors divided by number of predictions. MAE measures set of predicted value to actual value for instance, how close a predicted model is to the actual model.

3. **Root Mean Squared Error (RMSE):** Is defined as square root of sum of squares error divided number of predictions. RMSE measures the differences between values predicted by a model and the values actually observed. The smaller the value of RMSE means better accuracy of model.
4. **F-Measure**: This is a data mining statistic that measures both the Precision and recall of the model. It is used to provide a single measurement of model accuracy combining both precision and recall.
5. **ROC Area/Curve**: The Receiver Operating Characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. A receiver operating characteristics (ROC) curves is a technique for visualizing, organizing and selecting classifiers based on their performance.
6. **The Area Under ROC Curve (AUC)** frequently is used as a measure of the effectiveness of diagnostic markers.

A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system:

*0.90-1 = **excellent** (A), 0.80-0.90 = **good** (B), 0.70-0.80 = **fair** (C), 0.60-0.70 = **poor** (D), 0.50-0.60 = **fail** (F)*

A value near 0.5 means the lack of any statistical dependence.

ROC curves have an attractive property which is they are insensitive to changes in class distribution. If the proportion of positive to negative instances changes in a test set, the ROC curves does not change.

FIGURE 25
Decision Trees Evaluation Results

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          50               83.3333 %
Incorrectly Classified Instances        10               16.6667 %
Kappa statistic                          0.6656
Mean absolute error                      0.1973
Root mean squared error                  0.3628
Relative absolute error                 39.4625 %
Root relative squared error             72.509  %
Total Number of Instances               60

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.871     0.207      0.818     0.871     0.844      0.827    Yes
                 0.793     0.129      0.852     0.793     0.821      0.827    No
Weighted Avg.    0.833     0.169      0.834     0.833     0.833      0.827
```

FIGURE 26
PART Rule Evaluation Results

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          52               86.6667 %
Incorrectly Classified Instances         8               13.3333 %
Kappa statistic                          0.7324
Total Cost                               8
Average Cost                             0.1333
Mean absolute error                      0.1825
Root mean squared error                  0.3551
Relative absolute error                 36.4963 %
Root relative squared error             70.9767 %
Total Number of Instances               60

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.903     0.172      0.848     0.903     0.875      0.85     Yes
                 0.828     0.097      0.889     0.828     0.857      0.85     No
Weighted Avg.    0.867     0.136      0.868     0.867     0.866      0.85
```

FIGURE 27
Naive Bayes Evaluation Results

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          52               86.6667 %
Incorrectly Classified Instances         8               13.3333 %
Kappa statistic                          0.7324
Mean absolute error                      0.1759
Root mean squared error                  0.326
Relative absolute error                 35.1893 %
Root relative squared error             65.1644 %
Total Number of Instances               60

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.903     0.172      0.848     0.903     0.875      0.88     Yes
                 0.828     0.097      0.889     0.828     0.857      0.88     No
Weighted Avg.    0.867     0.136      0.868     0.867     0.866      0.88
```

TABLE 8
Algorithm Results Evaluation

Evaluation Result (Statistic Measure)	**Decision Tree (Algorithm I)**	**PART Rule (Algorithm II)**	**Naïve Bayes (Algorithm III)**
Kappa Statistic *highest value – Naïve Bayes*	0.6656	**0.7324**	**0.7324**
Mean absolute error *highest value – Naïve Bayes*	0.1973	0.1825	**0.1759**
Root mean square error *least value – Naïve Bayes*	0.3628	0.3551	**0.326**
Relative absolute error *Least value – Naïve Bayes*	39.4625%	36.4963%	**35.1893%**
F-Measure *highest value – Naïve Bayes*	0.833	**0.866**	**0.866**
ROC Area *highest value – Naïve Bayes*	0.827	0.85	**0.88**

The conclusion is that the inter-rater reliability is satisfactory, because the obtained Kappa of **0.7324** is higher than the commonly applied criteria of **0.70**. Kappa statistic varies from 0 to 1, in this case **0.60 – 0.79** = substantial agreement. The findings of the prediction model were accurate evidenced an F-Measure score of **0.866** for PART Rule and Naive Bayes algorithms. The Area Under ROC Curve (AUC) frequently is used as a measure of the effectiveness of diagnostic markers. The area under the curve was at **0.88**

FIGURE 28
Scatter Plot of P (LTFU=yes) vs Viral load

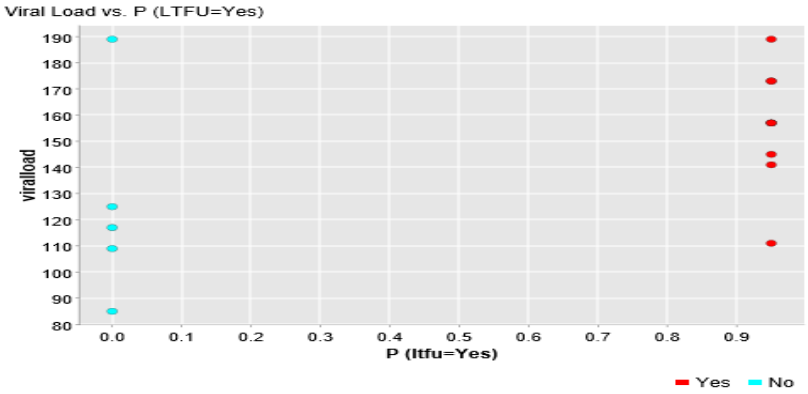


FIGURE 29
Scatter Plot of P (LTFU=yes) vs CD4 Count

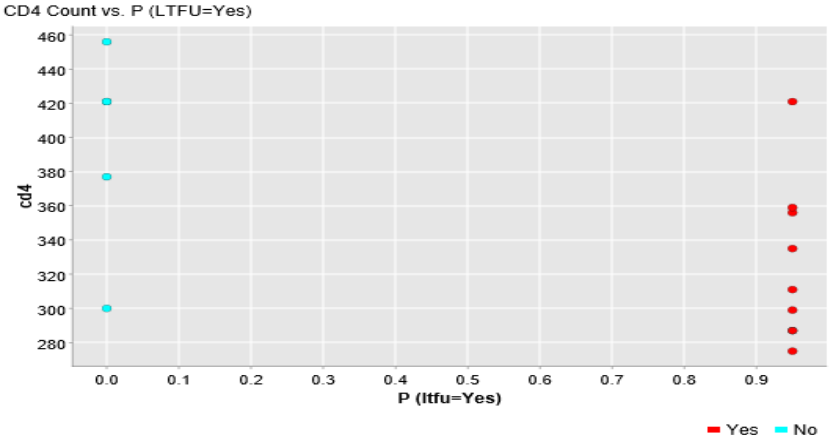


FIGURE 30
Scatter Plot of P (LTFU=yes) vs Window Period

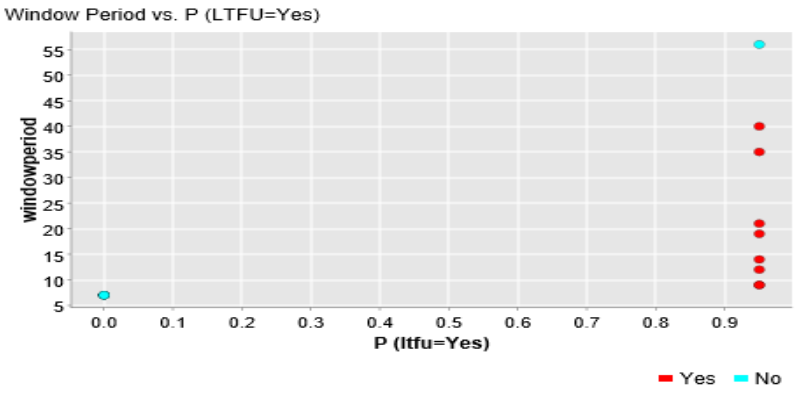


FIGURE 31
Scatter Plot of P (LTFU=yes) vs Age in years

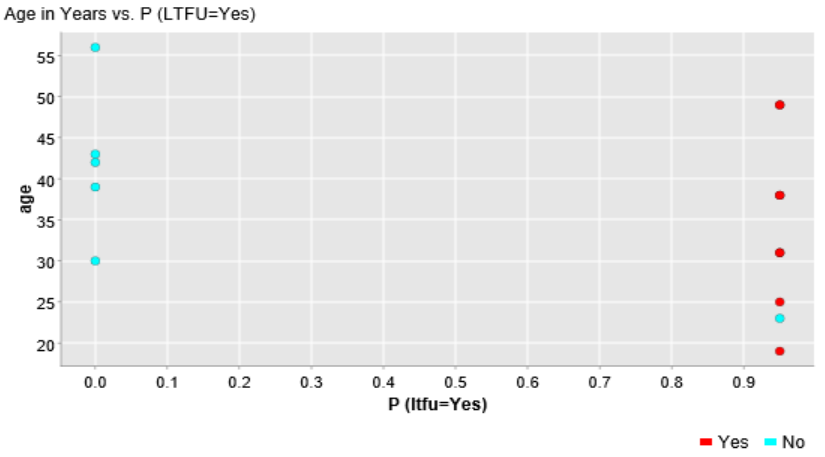
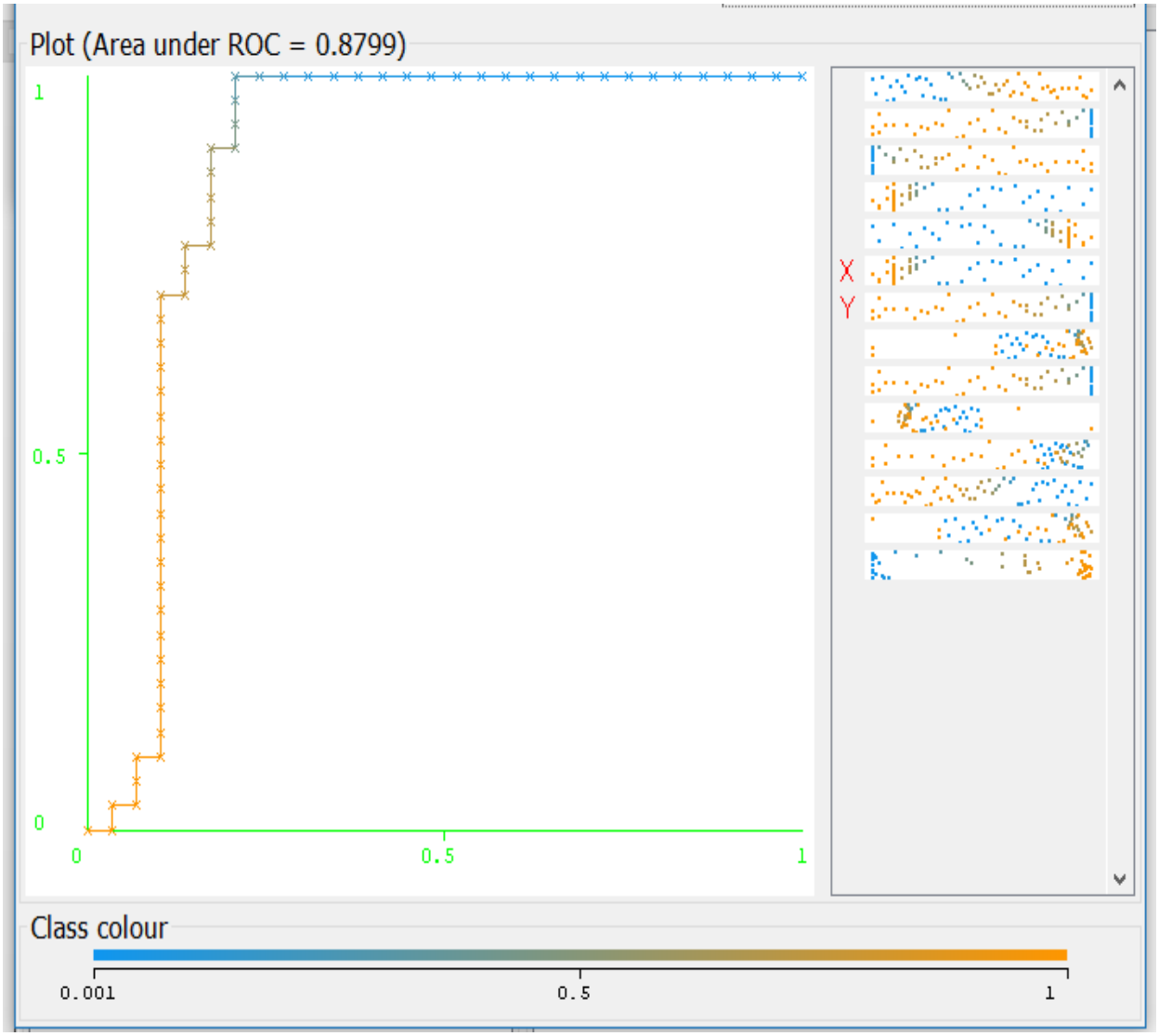


FIGURE 32
ROC Curve - Area Under ROC



ROC graphs are a very useful tool for visualizing and evaluating classifiers (LTFU predictors). They can provide a richer measure of classification performance than scalar measures such as accuracy, error rate or error cost. An efficient and efficacious use of the ROC graphs assesses classifiers and helps to promote better evaluation practices in the pattern recognition community.

CHAPTER FIVE
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents conclusions on the research findings and the recommendations drawn. that are to be used to further this work.

5.2 Summary

1. This study's findings are important as we try to develop a retention reinforcement strategy to address the multiple and complex challenges of patient retention in care.
2. Identifying the profile of individuals at risk is highly relevant as it enables development of interventions targeting patients at highest risk of LTFU.
3. Further efforts should be made to develop innovative strategies to identify and motivate individuals at highest risk of disengagement from care.
4. Analysis findings of the datasets analyzed suggest that LTFU tracing may lead to increased re-engagement of patients in care (PLHIVs).
5. The findings presented in this dissertation, along with others, can inform the development and implementation of strategies that encourage patients to remain in care and on treatment with lifesaving ART.

5.3 Conclusions

Bilateral and cross-border ART programs need to maintain current contact information of patients in care or their guardians, support buddies and should be aware of the limitations of tracing methods used. Emphasis is placed on their aiming for early prediction of LTFUs and tracing after missed appointment for ART pick-up, especially if mobile patients have recently started ART, including searching non/participating facility records for patients LTFU.

These interventions increase the proportion of successfully traceable patients through targeted retention and follow-up approaches to return them back to care (increase ARV drugs adherence), reduce mortality while improving overall program (bilateral and non-bilateral) performance reporting and evaluation.

5.4 Recommendations

1. Further research on ideal programming languages should be done to link the system's output to the data mining predictor model on (KNIME analytics) that mines the data, learns and predicts a possible LTFU, sending the systems analytics as a flag/color visualized on an analytics platform using python.
2. Scholars should further this study through exploring options of unique identification (national IDs, Names) of clients with available options (Biometrics) while ensuring confidentiality (stigma and discrimination) of the PLHIVs.
3. Security of PLHIVs data (encryption of sensitive data like identification information, HIV status, and locator information?) stored on a shared server, policy and governance, tool ownership and sustainability by local governments.

5.5 Contribution to knowledge

1. This study's findings are important as we try to develop a retention reinforcement strategy to address the multiple and complex challenges of patient retention in care.
2. Identifying the profile of individuals at risk is highly relevant as it enables development of interventions targeting patients at highest risk of LTFU.
3. Further efforts should be made to develop innovative strategies to identify and motivate individuals at highest risk of disengagement from care.
4. Analysis findings of the datasets analyzed suggest that LTFU tracing may lead to increased re-engagement of patients in care (PLHIVs).
5. The findings presented in this dissertation, along with others, can inform the development and implementation of strategies that encourage patients to remain in care and on treatment with lifesaving ART.

REFERENCES

AIDS Commission, U., 2009. UGANDA HIV Prevention Response and Modes of Transmission Analysis Uganda AIDS Commission. , (March).

Alvarez-Uria, G. et al., 2013. Factors associated with attrition, mortality, and loss to follow up after antiretroviral therapy initiation: Data from an HIV cohort study in India. *Global Health Action*, 6(1).

An AIDS free Uganda, M. responsibility! & Uganda AIDS Commission, R. of U., 2015. UAC (2015) National HIV and AIDS Strategic Plan 2015/2016- 2019/2020. , 2018, p.73. Available at: http://www.emtct-iatt.org/wp-content/uploads/2016/01/National-HIV-and-AIDS-strategic-plan-2015-16-2019-20.pdf.

AVERT, 2016. HIV and AIDS in Kenya | AVERT. *HIV AND AIDS IN KENYA*. Available at: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/kenya [Accessed July 22, 2016].

Berheto, T.M., Haile, D.B. & Mohammed, S., 2014. Predictors of loss to follow-up in patients living with hiv/aids after initiation of antiretroviral therapy. *North American Journal of Medical Sciences*, 6(9), pp.453–459.

Bygrave, H. et al., 2010. Trends in loss to follow-up among migrant workers on antiretroviral therapy in a community cohort in Lesotho. *PLoS ONE*, 5(10), pp.3–7.

CIA Fact book, 2015. Kenya vs. Uganda - demographics comparison. *Central Intelligence Agency Facts*. Available at: http://www.indexmundi.com/factbook/compare/kenya.uganda/demographics [Accessed February 16, 2017].

Cutsem, V. et al., 2012. Targeted adherence strategies for provision of cross-border antiretroviral therapy to migrant farm workers in Musina, South Africa. , 1.

Decroo, T. et al., 2011. Distribution of Antiretroviral Treatment Through Self-Forming Groups of Patients in Tete Province, Mozambique. *J Acquir Immune Defic Syndr*, 56(2), pp.e39–e44.

Dey, M. & Rautaray, S.S., 2014. Study and Analysis of Data mining Algorithms for Healthcare Decision Support System. *International Journal of Computer Science and Information Technologies*, 5(1), pp.470–477.

Eguzo, K.N. et al., 2015. Predictors of Loss to Follow-up Among HIV-infected Patients in a Rural South-Eastern Nigeria Hospital: A 5-year Retrospective Cohort Study. *Annals of medical and health sciences research*, 5(6), pp.373–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27057373 [Accessed June 10, 2017].

Fraser, H.S.F. et al., 2007. Information systems for patient follow-up and chronic management of HIV and tuberculosis: a life-saving technology in resource-poor areas. *Journal of medical Internet research*, 9(4), p.e29. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/17951213 [Accessed June 10, 2017].

Fuller, C.M., Ford, C. & Rudolph, A., 2009. *Injection drug use and HIV: Past and future considerations for HIV prevention and interventions*,

Gelmon, L. et al., 2009. Kenya: HIV prevention response and modes of transmission analysis. *Kenya National AIDS Control Council*, (March), p.83. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Prevention+Response+and+Modes+of+Transmission+Analysis#0.

Hailu, T.G., 2011. Comparing Data Mining Techniques in HIV Testing Prediction. *Intelligent Information Management Intelligent Informa-tion Management*, 7(7), pp.153–180. Available at: http://www.scirp.org/journal/iim%5Cnhttp://dx.doi.org/10.4236/iim.2015.73014%5Cnhttp://creativecommons.org/licenses/by/4.0/.

Hand, D. et al., 2001. *Principles of data mining*, Available at: http://www.ncbi.nlm.nih.gov/pubmed/17604416.

Harries, A.D. et al., 2010. Strategies to improve patient retention on antiretroviral therapy in sub-Saharan Africa. *Tropical Medicine and International Health*, 15(SUPPL. 1), pp.70–75.

Kaiser Family Foundation, 2016. The Global HIV / AIDS Epidemic: Fact Sheet. *The Henry Kaiser Family Foundation*, (July), p.4.

Kalechofsky, H. & Ph, D., 2016. A Simple Framework for Building Predictive Models A Little Data Science Business Guide Authored by : , (September).

Kerrigan, D. et al., 2003. Environmental-structural factors significantly associated with consistent condom use among female sex workers in the Dominican Republic. *AIDS (London, England)*, 17(3), pp.415–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12556696 [Accessed February 15, 2017].

Mberi, M.N. et al., 2015. Determinants of loss to follow-up in patients on antiretroviral treatment, South Africa, 2004-2012: a cohort study. *BMC health services research*, 15(July), p.259. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4491264&tool=pmcentrez&rendertype=abstract.

Megerso, A. et al., 2016. Predictors of loss to follow-up in antiretroviral treatment for adult patients in the Oromia region, Ethiopia. *HIV/AIDS (Auckland, N.Z.)*, 8, pp.83–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27175095 [Accessed June 10, 2017].

Meloni, S.T. et al., 2014. Time-dependent predictors of loss to follow-up in a large HIV treatment cohort in Nigeria. *Open Forum Infectious Diseases*, 1(2), pp.18–23.

Murphy, K. (Kenneth M.. et al., 2008. *Janeway's immunobiology.*, Garland Science.

NACC, 2014. Kenya AIDS Strategic Framework 2014/2015 - 2018/2019. *National AIDS Control Council*, pp.1–84. Available at:

http://www.nacc.or.ke/attachments/article/181/KNASP III.pdf.

NACC, 2008. United Nations General Assembly Special Session on HIV and AIDS: Country report - Kenya.

NACC and NASCOP, 2012. National strategy on hiv and aids & sti programming along transport corridors in kenya.

NASCOP, 2014. Kenya AIDS Indicator Survey 2012: Final Report. *Journal of acquired immune deficiency syndromes (1999)*, 66 Suppl 1, pp.1–530. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24732813.

National AIDS Control Council, 2014. Kenya AIDS Response Progress Report Progress towards Zero. *Nascop, Ministry of Health, Government of Kenya.*, (March), p.37.

ONUSIDA, 2013. The Gap Report. , (1), pp.1–12.

Oo, M.M. et al., 2016. Alarming attrition rates among HIV-infected individuals in pre-antiretroviral therapy care in Myanmar, 2011-2014. *Global health action*, 9, p.31280. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27562473 [Accessed June 9, 2017].

Tadesse, K. & Haile, F., 2014. Predictors of Loss to Follow Up of Patients Enrolled on Antiretroviral Therapy : *J AIDS Clin Res*, 5(12).

Tang, W. et al., 2015. Factors associated with loss-to-follow-up during behavioral interventions and HIV testing cohort among men who have sex with men in Nanjing, China. *PLoS ONE*, 10(1).

The National ART Program Swaziland, 2012. PATIENT LINKAGE , RETENTION AND FOLLOW-UP in HIV CARE STANDARD OPERATING PROCEDURES. , (February).

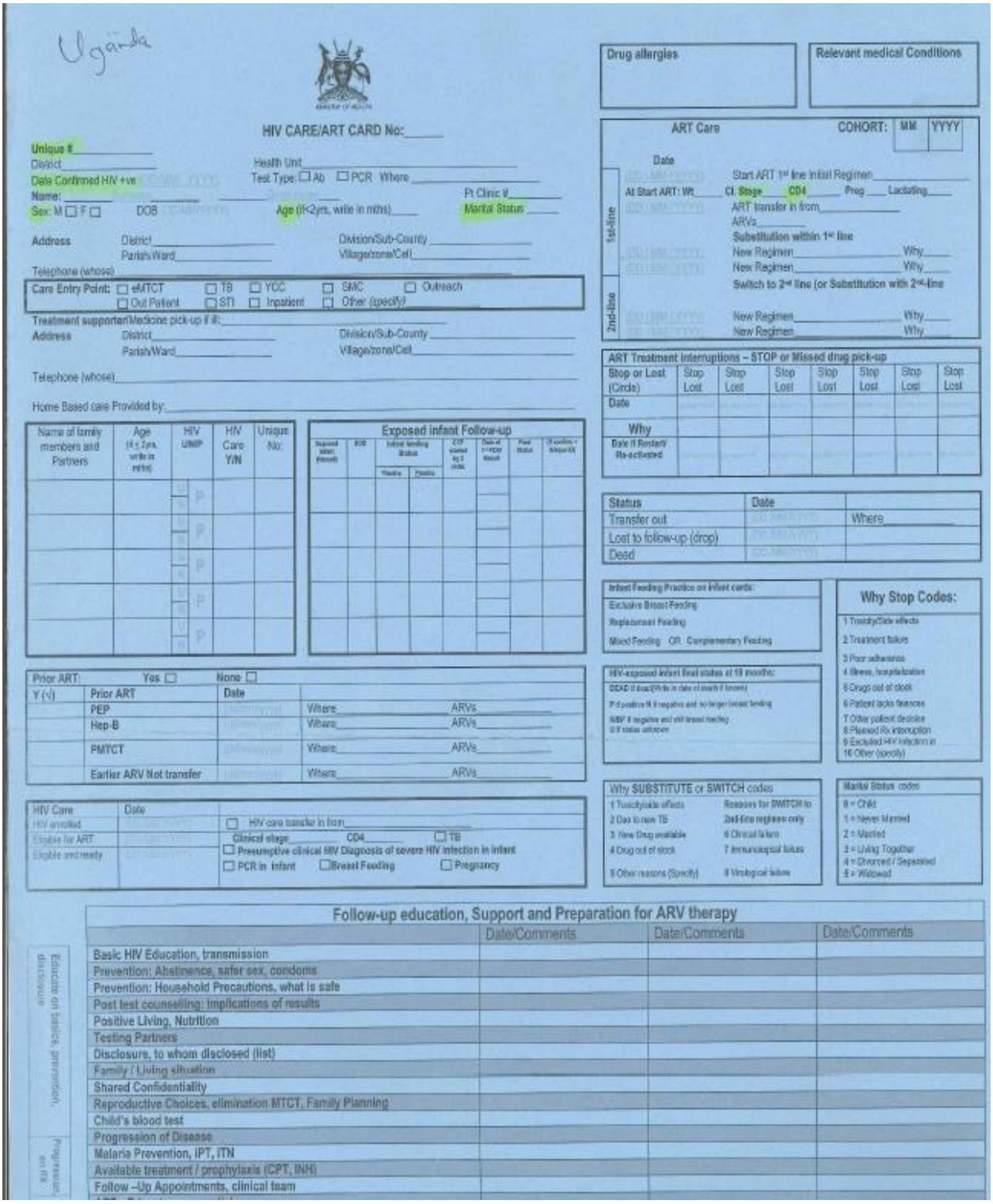
Tiruneh, Y.M. et al., 2016. Retention in Care among HIV-Infected Adults in Ethiopia, 2005-2011: A Mixed-Methods Study. *PloS one*, 11(6), p.e0156619.

UNAIDS, 2010. GOVERNMENT OF UGANDA UNGASS COUNTRY PROGRESS REPORT UGANDA January 2008-December 2009 March 2010. *Development*, (March).

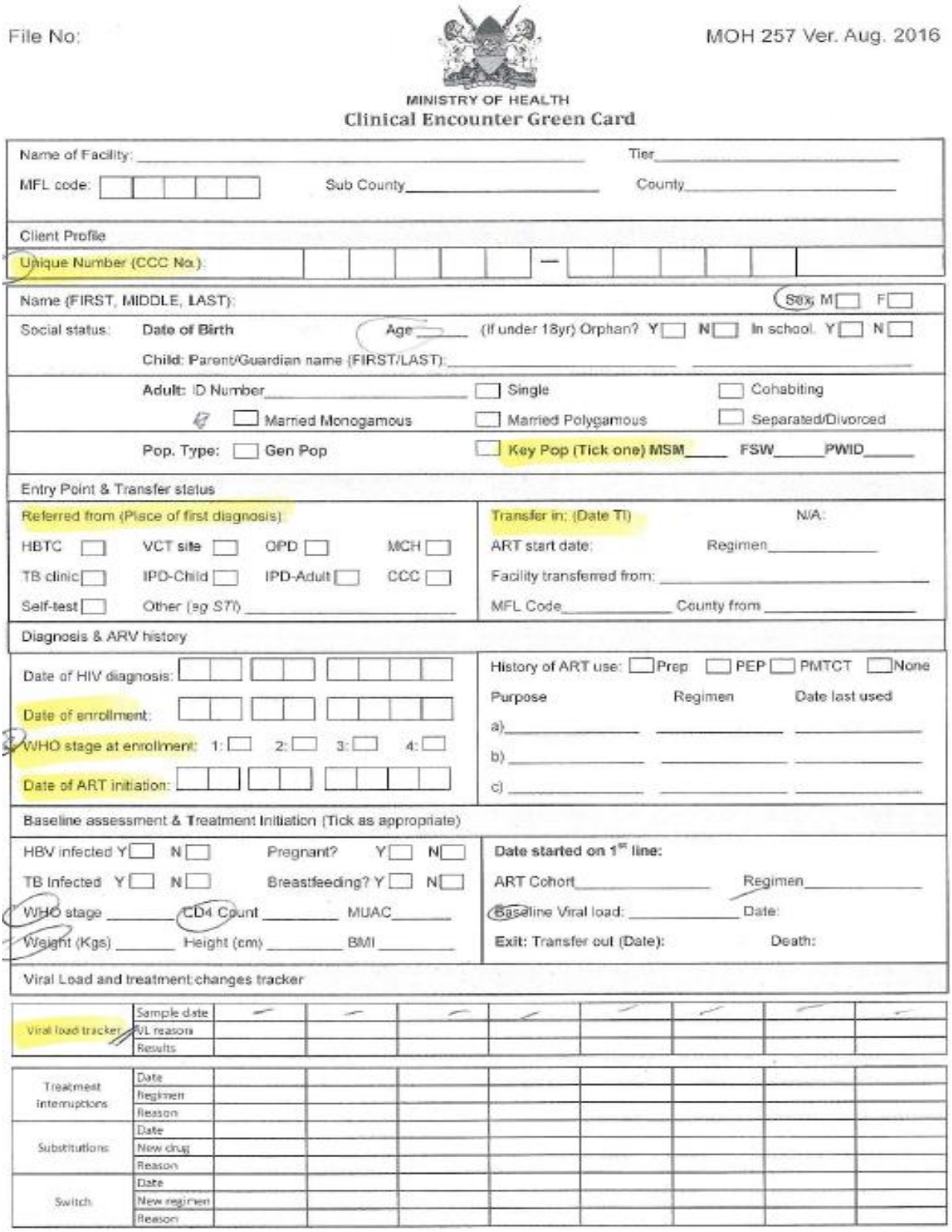
UNAIDS, 2016. Unaids Gap Report 2016.

APPENDICIES

APPENDIX I: FIGURE A1 - Uganda Daily Activity Register (DAR)



APPENDIX II: FIGURE B1 - Kenya Daily Activity Register (DAR)



File No:

MOH 257 Ver. Aug. 2016

MINISTRY OF HEALTH

Clinical Encounter Green Card

Name of Facility: _____ Tier _____

MFL code: ☐☐☐☐☐ Sub County _____ County _____

Client Profile

Unique Number (CCC No.): ☐☐☐☐☐ – ☐☐☐☐☐☐

Name (FIRST, MIDDLE, LAST): Sex: M ☐ F ☐

Social status: Date of Birth Age ____ (If under 18yr) Orphan? Y ☐ N ☐ In school. Y ☐ N ☐

Child: Parent/Guardian name (FIRST/LAST): _____ _____

Adult: ID Number _____ ☐ Single ☐ Cohabiting

☐ Married Monogamous ☐ Married Polygamous ☐ Separated/Divorced

Pop. Type: ☐ Gen Pop ☐ Key Pop (Tick one) MSM____ FSW____ PWID____

Entry Point & Transfer status

Referred from (Place of first diagnosis)

HBTC ☐ VCT site ☐ OPD ☐ MCH ☐

TB clinic ☐ IPD-Child ☐ IPD-Adult ☐ CCC ☐

Self-test ☐ Other (eg STI) _____

Transfer in: (Date TI) N/A:

ART start date: Regimen _____

Facility transferred from: _____

MFL Code _____ County from _____

Diagnosis & ARV history

Date of HIV diagnosis: ☐☐ ☐☐ ☐☐☐☐

Date of enrollment: ☐☐ ☐☐ ☐☐☐☐

WHO stage at enrollment: 1: ☐ 2: ☐ 3: ☐ 4: ☐

Date of ART initiation: ☐☐ ☐☐ ☐☐☐☐

History of ART use: ☐ Prep ☐ PEP ☐ PMTCT ☐ None

Purpose	Regimen	Date last used
a)		
b)		
c)		

Baseline assessment & Treatment Initiation (Tick as appropriate)

HBV infected Y ☐ N ☐ Pregnant? Y ☐ N ☐

TB Infected Y ☐ N ☐ Breastfeeding? Y ☐ N ☐

WHO stage _____ CD4 Count _____ MUAC _____

Weight (Kgs) _____ Height (cm) _____ BMI _____

Date started on 1st line:

ART Cohort _____ Regimen _____

Baseline Viral load: _____ Date:

Exit: Transfer out (Date): Death:

Viral Load and treatment changes tracker

Viral load tracker	Sample date								
	VL reason								
	Results								
Treatment interruptions	Date								
	Regimen								
	Reason								
Substitutions	Date								
	New drug								
	Reason								
Switch	Date								
	New regimen								
	Reason								

APPENDIX III: FIGURE C1 - Harmonized inter-facility referral form

Section A: Client personal identification *Retained at facility*

Client Registration No.		Date of referral:	
Full Names:		Date of birth:	
Sex (M/F):		Age in years:	
Nationality:		Marital Status:	
Referring Facility Name:		Country:	
Referred facility Name:		Country:	
District/County:		CB-HIPP Site [1]:	
Type of care*:		Target Population [2]:	
Contact/address		Support / treatment buddy:	
		Contact / address:	

Section B: Reason for referral

1. TB/HIV: ☐ 2. ART: ☐ 3. TB ☐ 4. STI: ☐ 5. FP/RH: ☐ 6.MNCH/ANC: ☐ 7.GBV: ☐ 8. Other: ☐ Please specify: _____

Description/specific details (see below list)

Description details*:

1. Where tested (Name of facility, county/district and country
2 **HIV**: Date confirmed HIV positive: Weight, CD4, VL at baseline & referral; TB stage, Type, Sputum; Treatment - Start date, Drug/medication: Dose, regime, 1st, 2nd, 3rd line, if on ART IPT, WHO clinical stage, status e.g. Bed ridden?
3. **TB:** TB stage, Type (PTB, EPTB, DRTB), genexpert, Sputum; IPT, WHO clinical stage, medication received, regime, 1st, 2nd, 3rd line, Treatment - Start date, Drug: Dose, if on ART?
4. **STI**: Drugs: Dosage: Start date; Type/syndrome, ug/medication: Dose, regime, 1st, 2nd, 3rd line,
5. **FP/RH**: Preg, status, Method last used: Method currently using, not applicable e.g. pregnant, male, post-menopausal women, FP method received
6. **MNCH/ANC:** Trimester, Date of first ANC visit, 1st 2nd, 3rd, 4th HIV test/status, PMTCT, LMP & EDD, vaccination, postnatal care, other
7. **GBV**: Incident (extortion, abuse, arrest, assault, sexual harassment, rape etc.): Service (legal, PEP, Psychosocial)
8. **ART**: Enrollment date, regimen, Drugs, Start date, 1st, 2nd or 3rd line
9. **Other** Type of care, please specify

General comments if referred for ART:		***Referral Recommendations:***	
Hypersensitivity:		Continue CTX and Follow up:	
Drug Toxicity:		Eligible for ART preparation:	
Poor Adherence:		Continue ART:	

Section C: Referring Health Care Provider

Full Names:		Signature:	
Email/Tel with Country Code:		Cadre:	

Section D: Receiving Health Care Provider

Full Names:		Signature:	
Email/Tel with Country Code:		Cadre:	
Receiving Facility Name:			
Action Taken:			

Section E: Feedback

LEGEND:

Site: 1. Busia 2. Malaba 3. Sio Port 4. Port Victoria 5. Taveta 6. Majanji 7. Katuna 8. Kasensero 9. Holili 10. Rubavu, 11. Rusumo 12. Gatuna 13. Kirongwe 14. Kabonga 15. Muhuru

Target Population: (1). FSW (2). Truckers (3). Freight Clearing and Forwarding Agents (4). Fisher folk (5). MSM (6). VWG (7). PWID (8). PLHIV and discordant couples (9). Other (specify)

Acronyms: HTS - HIV testing Service, ART - Antiretroviral Therapy, STI - Sexually Transmitted Infections, TB - Tuberculosis, MNCH - maternal Child Health, FP/RH - Family planning, GBV - Gender Based Violence