# EVALUATION OF A HIV PREDICTIVE ALGORITHM, GEOSPATIAL ANALYSIS OF NEW HIV DIAGNOSES, AND MAPPING OF HIV TESTING UPTAKE IN HOMA BAY, KISUMU, AND SIAYA COUNTIES, WESTERN KENYA

**BY**

**HELLEN MUTTAI**

**A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PUBLIC HEALTH**

**SCHOOL OF PUBLIC HEALTH AND COMMUNITY DEVELOPMENT**

**MASENO UNIVERSITY**

## DECLARATION

This thesis is my original work and has not been presented for the award of a degree in any other university.

**Student:**

**Hellen Muttai**

PH/PHD/00023/2017

Sign……  …… Date…6 November 2023……

**Declaration by Supervisors:**

This thesis has been submitted for examination with our approval as University Supervisors.

**Dr. Bernard Guyah, PhD**

School of Public Health and Community Development,

Maseno University, Maseno, Kenya

Sign……..…… Date…6 November 2023……

**Dr. Emily Zielinski-Gutierrez, PhD**

United States Centers for Disease Control and Prevention,

Guatemala, Central America

Sign……  ……..…… Date…6 November 2023……

## ACKNOWLEDGEMENT

I am grateful for the mentorship that I have received from my supervisors: Dr. Bernard Guyah and Dr. Emily Zielinski-Gutierrez. I thank Professor Thomas Achia and Paul Musingila for their statistical mentorship and support for this study; and the contributions of the co-authors of the manuscripts published.

I thank the Kenya Ministry of Health and the County Departments of Health for Kisumu, Homa Bay, and Siaya counties, for continuing to provide HIV services. I express my gratitude to the staff who implemented routine provider-initiated testing and counseling at the health facilities described in this study and home-based HIV testing in Siaya County, without whom this study would not have been possible.

Last, I thank my family for their support throughout my doctoral studies.

## DEDICATION

I dedicate this thesis to my parents who gave me the opportunity for a good education and instilled in me the passion to work hard; and to my family, especially my three sons, who have inspired me to greater heights.

## ABSTRACT

As the number of people living with HIV in the population who do not know their HIV status continues to decline, as more people are linked to ART, continuing to offer HIV testing in a universal manner becomes inefficient. Finding ways to target HIV testing to persons more likely to be HIV positive, for efficiency, is a global priority. The purpose of this study was to evaluate the use of three strategies to identify sub-populations and granular-geographic areas with higher HIV positive yield to inform efficient targeting of HIV testing among persons ≥15 years in Homa Bay, Siaya and Kisumu; counties with the highest HIV prevalence and incidence in Kenya. The specific objectives were to evaluate the use of a HIV predictive risk-score algorithm, geospatial analysis of new HIV diagnoses, and mapping of HIV testing uptake. Using a hospital-based retrospective cohort study design, a HIV predictive risk-score screening algorithm was developed using univariable and multivariable analyses of outpatient data, comprising 19,458 persons ≥15 years tested for HIV from September 2017–May 2018 from five purposively selected health facilities in Homa Bay, Siaya and Kisumu Counties. Using a community-based retrospective cohort study design, the use of geospatial analysis to assess geospatial patterns of new HIV diagnoses, and the use of mapping HIV testing uptake, were evaluated. Community home-based data comprised 365,798 clients aged ≥15 years offered home-based HIV testing as part of a routine public health program from May 2016–July 2017 in Siaya County. Geospatial analysis using Kulldorff's spatial scan statistic was used to detect geographic clusters (radius <5 kilometers) of new HIV diagnoses. A Geographical Information System program was used to map HIV testing uptake. The results showed that an HIV predictive risk-score screening algorithm developed grouped patients into four risk-score categories: <9, 10–15, 16–29 and >30, with increasing HIV prevalence of 0.6% [95% Confidence Interval (CI): 0.46–0.75], 1.35% (95% CI: 0.85–1.84), 2.65% (95% CI: 1.8–3.51), and 15.15% (95% CI: 9.03–21.27), respectively. External validation of the algorithm produced similar results. The algorithm's discrimination performance was modest, with an area under the receiver-operating-curve of 0.69 (95% CI: 0.53–0.84). The algorithm accounted for a high proportion ($R^2$ 0.89) of the variability of HIV prevalence in the study population. Results from geospatial analysis of new HIV diagnoses showed spatial variation in the distribution of new HIV diagnoses, and nine sub-location clusters in which the number of new HIV diagnoses was significantly (1.56 to 2.64 times) higher than expected were identified. Results from mapping HIV testing uptake found that 268,543 (86%) clients were tested for HIV. Of the 43,680 eligible clients not tested, 32,852 (75%) were not found at home and 5,931 (14%) declined testing. Granular geographic areas with low testing uptake, a high proportion of clients not found at home and a high proportion who declined testing, yet with clusters of higher new HIV diagnoses were identified. In conclusion, the following strategies successfully identified sub-populations and granular-geographic areas with higher HIV positive yield that should be targeted in the implementation of HIV testing services: a HIV predictive risk-score screening algorithm that identified patients who are more likely to be HIV-positive; geospatial analysis that identified granular sub-location clusters (<5 kilometers) of higher new HIV diagnoses; and mapping of HIV testing uptake that identified granular-geographic areas with low HIV testing uptake yet higher HIV positive yield. These study findings inform global, national, and county government policies and strategies for targeting HIV testing, for efficient use of resources and maximal epidemiologic impact.

## TABLE OF CONTENT

## LIST OF ABBREVIATIONS AND ACRONYMS

| | | |
|---|---|---|
| **AIDS** | : | Acquired Immunodeficiency Syndrome |
| **ANOVA** | : | Analysis of Variance |
| **AOR** | : | Adjusted Odds Ratio |
| **ART** | : | Antiretroviral Therapy |
| **AUC** | : | Area under the Receiver Operating Curve |
| **BYM** | : | Besag-York-Mollié |
| **CAR** | : | Conditional Autoregressive |
| **CARET** | : | Classification and REgression Training |
| **CD4** | : | Cluster of differentiation 4 |
| **CDC** | : | Centers for Disease Control and Prevention |
| **CI** | : | Confidence Interval |
| **CrI** | : | Credible Interval |
| **DREAMS** | : | Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe women |
| **EPV** | : | Event per Variable |
| **FSW** | : | Female Sex Workers |
| **GIS** | : | Geographic Information System |
| **HIV** | : | Human Immunodeficiency Virus |
| **HP Cluster** | : | High Prevalence Cluster |
| **HPTN** | : | HIV Prevention Trials Network |
| **IMCI** | : | Integrated Management of Childhood Illness |
| **INLA** | : | Integrated Nested Laplace Approximation |
| **IPD** | : | In-patient Department |

| | | |
|---|---|---|
| **IQR** | : | Interquartile Range |
| **Km** | : | Kilometers |
| **MOH** | : | Ministry of Health |
| **MSM** | : | Men who have Sex with Men |
| **OPD** | : | Out-patient Department |
| **OR** | : | Unadjusted Odds Ratio |
| **PEPFAR** | : | President's Emergency Plan for AIDS Relief |
| **PITC** | : | Provider Initiated Testing and Counseling |
| **PLHIV** | : | People Living with HIV |
| **PrEP** | : | Pre-exposure prophylaxis for HIV |
| **PWID** | : | People who Inject Drugs |
| **QGIS** | : | Quantum Geographic Information System |
| **RE&IC** | : | Risk of Exposure and Indicator Conditions |
| **$R^2$** | : | R-squared |
| **RR** | : | Relative Risk |
| **SAS** | : | Statistical Analysis System |
| **STI** | : | Sexually Transmitted Infections |
| **TB** | : | Tuberculosis |
| **UNAIDS** | : | United Nations Joint Programmed on HIV/AIDS |
| **USA** | : | United States of America |
| **WHO** | : | World Health Organization |

## DEFINITION OF TERMS

**Concentrated HIV epidemic**- HIV has spread rapidly in a defined sub-population (such as men who have sex with men, sex workers, transgender people, people who use drugs or people in prison or other closed settings), but is not well established in the general population.

**Early antiretroviral therapy (ART)** - is initiation of ART soon after HIV diagnosis, regardless of the immunologic or clinical status of an individual.

**Emergency department-** is a medical treatment unit that offers emergency medicine, the acute care of patients who present without prior appointment; either by their own means or by that of an ambulance. The emergency department is usually found in a health facility, hospital.

**Generalized HIV epidemic**- is where HIV is firmly established in the general population. Although sub-populations at high risk may contribute disproportionately to the spread of HIV, sexual networking in the general population is sufficient to sustain the epidemic.

**Geospatial analysis-** is the processing or manipulation of data that has a geographic or spatial component to identify patterns.

**Granular geographic mapping-** refers to mapping of geospatial data to small geographic units, for this study sub-location level.

**Home-based HIV testing-** a HIV testing strategy, where counselors move from one house to another, enumerating people who live in each household and providing HIV testing.

**HIV positive yield**- proportion of HIV positive individuals identified from those tested for HIV.

**Implementation science**- refers to the scientific study of methods to promote the systematic uptake of research findings or other evidence-based practice into routine health care to improve the quality and effectiveness of health services.

**Key populations**- include female sex workers (FSWs), men who have sex with men (MSM), and people who inject drugs for pleasure (PWID).

**Opt-out approach for HIV testing**- offering HIV testing routinely to all clients as standard of care, unless they decline.

**Outpatient department-** is the part of a health facility or hospital designed for the treatment of outpatients, people with health problems who visit the facility or hospital for diagnosis or treatment, but do not at this time require a bed or to be admitted for overnight care.

**Predictive algorithm**- a set of multiple variables put together by combining individual variables, that is statistically used to show prediction (the likelihood of happening) of an outcome.

**Priority populations**- includes the fishing community.

**Proportion of new HIV positive clients (new HIV positive yield)**- the total number of clients newly identified HIV positive among those with a conclusive test result in home-based HIV testing.

**Proportion of total HIV positive clients**- was calculated as the sum of new HIV positive and previously identified HIV-infected clients among those assessed for HIV test eligibility in home-based HIV testing.

**Strategy**- refers to a plan or set of actions designed to achieve a certain goal.

**Viral load**- the amount of HIV virus in plasma, measured through ribonucleic acid testing.

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER ONE
# INTRODUCTION

## 1.1 Background of the study

Globally, in 2019, there were about 38 million people living with Human Immunodeficiency Virus (HIV) (Joint United Nations Programme on HIV/AIDS, 2020b). Access to life-saving antiretroviral therapy (ART) has rapidly expanded in the past decade; by 2019, 67% (~25.4 million) of people living with HIV globally were accessing antiretroviral therapy, an increase from 7.8 million in 2010 (Joint United Nations Programme on HIV/AIDS, 2020b). Despite this, HIV continues to be a global public health threat. Although worldwide progress has been made in reducing new HIV infections among children, with new pediatric infections declining by 52% since 2010, there has been a slow decline in new infections among adults, adult new infections only reducing by 17%, from 1.8 million in 2010 to 1.5 million in 2019 (Joint United Nations Programme on HIV/AIDS, 2020b). Increased access to antiretroviral therapy has averted an estimated 12.1 million HIV-related deaths since 2010. In spite of this progress, hundreds of thousands of people (about 690,000 in 2019), are dying each year of a disease that has multiple effective and relatively inexpensive treatment regimens available (Joint United Nations Programme on HIV/AIDS, 2020b).

The sub-Saharan Africa region bears the brunt of HIV infection globally. In 2019, 25.6 million people were living with HIV in this region (67% of the global burden) (Joint United Nations Programme on HIV/AIDS, 2020b). The region has had rapid scale-up in HIV treatment; by 2019, 72% (18 million) of people living with HIV were accessing ART. Despite this, the sub-Saharan region contributes the highest number of global new HIV infections and deaths: in 2019, about 970,000 people were newly infected with HIV, accounting for 57% of new infections

globally; and 440,000 deaths occurred due to HIV-related causes, accounting for 64% of deaths globally (Joint United Nations Programme on HIV/AIDS, 2020b).

Kenya has an adult HIV prevalence of 4.2% and HIV incidence of 1.2 per 1000 population. In 2019, nationally, an estimated 1.5 million people were living with HIV. Kenya has made progress in increasing ART coverage; by September 2020, the country had achieved an ART coverage of 79%, with over 1.19 million people accessing ART (President's Emergency Plan for AIDS Relief, 2020). Although new infections have reduced by almost 50% in the last decade, from ~73,000 in 2010 to about ~40,000 in 2019, new infections continue to be high (Kenya National AIDS and STI Control Programme, 2020). Additionally, in 2019 about 21,000 people died of HIV-related causes (Joint United Nations Programme on HIV/AIDS, 2020b), which would have been preventable with effective ART.

The counties of Homa Bay, Siaya and Kisumu in the western region of Kenya have the highest adult HIV prevalence in Kenya, ranging from 14% to 18%; and incidence, ranging from 5.1 to 6.7 per 1000 population (Kenya National AIDS and STI Control Programme, 2020). These three counties have a total population of about 3 million people and about 372,000 people living with HIV; and by September 2020,318,903 people were on ART, achieving 85% ART coverage (President's Emergency Plan for AIDS Relief, 2020). Additionally, these three counties have an estimated 10,000 annual new infections, accounting for about 25% of new HIV infections in Kenya (Kenya National AIDS and STI Control Programme, 2020).

Since HIV continues to be a major public health threat, controlling the HIV epidemic, in order to realize both a public health impact by significantly reducing HIV-related morbidity and mortality, and an economic impact through future significant cost-savings, is a global priority.

Early ART with viral suppression has been shown to significantly reduce new HIV infections; it has been demonstrated to have an efficacy of >93% in preventing HIV transmission, the highest demonstrated among multiple prevention interventions that have been evaluated (Cohen et al., 2016; Dieffenbach, 2012). Additionally, early ART significantly reduces HIV-related morbidity and mortality (Cohen et al., 2011; Group., 2015). Based on this, in order to control the HIV epidemic, the Joint United Nations Programme on HIV/AIDS (UNAIDS) set ambitious global 90-90-90 targets to be achieved by 2020, recommending that programs aim for 90% of all people living with HIV to know their HIV status, 90% of all people with diagnosed HIV infection to receive sustained ART, and 90% of all people receiving ART to achieve viral suppression (Joint United Nations Programme on HIV/AIDS, 2014). Modeling suggests that achieving these treatment targets will enable the world to control the HIV epidemic, which in turn will generate profound health and economic benefits.

By the 2020 timeline set by UNAIDS to achieve the 90-90-90 target, the world and the sub-Sahara African regions were far from achieving this target. In 2020, the global ART coverage was 68%, and in sub-Sahara Africa 72%. Additionally, Kenya had achieved an ART coverage of 79%, and the three counties of Homa Bay, Kisumu and Siaya, 85%; although higher than the average for the sub-Saharan Africa region, Kenya and the three counties also fell short of achieving the 90-90-90 target. In December 2020, to further enhance achievement of HIV epidemic control, UNAIDS launched a new set of 95-95-95 target (Joint United Nations Programme on HIV/AIDS, 2020a), recommending that programs aim for 95% of all people living with HIV to know their HIV status, 95% of all people with diagnosed HIV infection to receive sustained ART, and 95% of all people receiving ART to achieve viral suppression.

HIV testing services is the entry point to identifying individuals who are HIV positive to link them to ART, in order to increase ART coverage and achieve HIV epidemic control. Implementation of HIV testing services at the country-level are guided by both global (WHO) and national policies and guidelines. The 2015 World Health Organization (WHO) guidelines for HIV testing (World Health Organization, 2015) and the Kenya HIV testing guidelines (Kenya Ministry of Health, 2015) recommended offering routine HIV testing to all clients attending health facilities, through a universal approach, with annual retesting or more frequently based on HIV-exposure or risk; and a strategic mix of community-based testing services.

Following these guidelines, in the past 8 years (since the launch of the 2015 HIV testing guidelines), Kenya has implemented multiple HIV testing strategies, that include universal testing of clients attending health facility services, through provided initiated testing and counseling (i.e. testing in outpatient and inpatient departments, maternal-child-health clinics, tuberculosis clinics, and index testing), and testing in community settings (i.e. mobile outreaches, specific settings for adolescent young women and men, voluntary medical circumcision clinics, home-based testing and out-reach index testing). As a result, in 2019, Kenya tested a total of 10,182,944 people, and identified 175,858 who were HIV positive, resulting in a HIV positive yield of 1.7% (President's Emergency Plan for AIDS Relief, 2020). Similarly, the three counties of Homa Bay, Kisumu and Siaya, in 2019 tested a total of 2,056,023 people, to identify 40,546 HIV positive, resulting in a HIV positive yield of 2.0% (President's Emergency Plan for AIDS Relief, 2020). Based on this data, although many clients were tested for HIV, a low proportion of those HIV positive were identified. This is mainly because as more people are identified and linked to ART, fewer people living with HIV remain undiagnosed in the population that need to be identified. Consequently, continuing to implement universal HIV testing to all clients

attending health facilities, as recommended by the 2015 WHO and Kenya HIV testing guidelines, and generalized community-based testing, has led to testing a lot of people, yet identifying a low proportion of HIV positive individuals (President's Emergency Plan for AIDS Relief, 2020). To be more efficient in HIV testing, strategies that tease out and identify sub-populations or geographic units that are more likely to yield a higher proportion of HIV positive individuals, that then would be targeted and offered HIV testing, are needed. This would lead to only testing people who are more likely to be HIV positive, or geographic units more likely to have a higher HIV positive yield, hence overall achieving efficiency by testing fewer people (unlike universal testing), and identifying a higher proportion of those HIV positive.

This study evaluated three strategies used to tease out and identify sub-populations and granular-geographic areas that have higher HIV positive yield, to inform targeting of HIV testing among persons ≥15 years of age, for testing efficiency: a HIV predictive risk-score screening algorithm; geospatial analysis of new HIV diagnoses; and granular-level mapping of HIV testing uptake.

### 1.2 Statement of the problem

As highlighted in the background, although progress has been made towards achieving HIV epidemic control, HIV continues to be a global public health threat. Achieving the 90-90-90 UNAIDS target, reset to 95-95-95, is a global priority for HIV epidemic control. Although UNAIDS aimed that the 90-90-90 target would be achieved by 2020, globally, in sub-Sahara Africa, and in Kenya, the target was not met. HIV testing is the entry point to increasing ART coverage and achieving the 90-90-90 target. As countries make progress in increasing ART coverage, a high proportion of people living with HIV are identified and linked to ART; which means fewer people living with HIV remain undiagnosed in the population and need to be identified. Offering HIV testing in a universal manner to the leads to many clients being tested

for HIV, and a low proportion of those HIV positive identified, which is not efficient. This study explored ways that can be used to tease out and identify sub-populations and geographic units that have higher HIV positive yield, to inform targeting of HIV testing for efficiency.

To identify sub-populations with higher HIV positive yield, this study developed and validated a HIV predictive risk-score algorithm. A predictive algorithm in this study refers to a set of variables or patient characteristics, when combined, is used to identify clients with a certain outcome, for this study, those more likely to be HIV positive. Several studies have evaluated HIV testing screening algorithms among children, adolescents, key populations, and women. Although many adult patients around the world flow through health facility outpatient departments to seek health services, few HIV screening algorithms are available or have been evaluated among adults in the facility-outpatient setting. Studies evaluating screening algorithms among adult outpatient attendees have been conducted in the United States and Spain. These are settings of low HIV prevalence and concentrated HIV epidemics, where HIV transmission largely occurs in defined sub-populations, mostly among key populations. During this study's literature review, no HIV screening algorithms or studies were found that have evaluated algorithms among adults in the outpatient setting in sub-Sahara Africa, where the HIV epidemic is generalized. A generalized epidemic is where HIV is firmly established in the general population, and transmission is largely driven by general population heterosexual networks. Therefore, HIV-risk factors in generalized epidemics are mostly related to heterosexual relationships in the general population, which differs from concentrated epidemics, where risk factors are mostly related to sexual relations within affected sub-populations. Developing a screening algorithm for use in the outpatient department, that is context specific to the sub-Saharan Africa region setting of a generalized epidemic, was noted as a major gap.

Geospatial analysis is another potential way to identify geographic areas with higher HIV positive yield to target HIV testing services for efficiency. Geospatial analysis is the processing or manipulation of data that has a geographic or spatial component to identify patterns. Geospatial analysis has been widely used to demonstrate geospatial variation and clustering of HIV infection around geographic, social, or behavioral risk factors. Furthermore, multiple studies have described ways to prioritize HIV interventions to specific geographic areas, including areas with higher HIV prevalence, higher HIV incidence, and focused prioritization based on local epidemiologic context. Despite this, few studies have spatially described or mapped new HIV diagnoses. A study conducted in Kenya mapped new HIV diagnoses using routine facility-level HIV testing data to 50-kilometer radius areas, across counties with differing HIV burden (Waruru et al., 2021b). Geospatial analysis and mapping of new HIV diagnoses to smaller geographic units was noted as a major gap, and would be programmatically useful for more granular targeting of HIV interventions. This study explored geospatial analysis of new HIV diagnoses to the smallest possible geographic unit. Mapping to village level was desired; however, it was not statistically feasible, as the total population of the village, the number of clients tested for HIV, and those identified as HIV positive were too small for meaningful statistical analysis. The next level of administrative unit, that was statistically feasible, was the sub-location unit. By using sub-location units, mapping of clusters of new HIV diagnoses was done to granular 5-kilometer radius areas. This study therefore uniquely conducted geospatial analysis at the smallest statistically feasible geographic unit, to inform granular targeting of HIV testing for efficiency.

Yet another potential strategy to identify geographic areas to target HIV testing that this study explored was mapping HIV testing uptake. Many studies have described home-based HIV

testing programs in sub-Sahara Africa and in Kenya. Geographic variation in HIV testing uptake has mostly been described for large geographic units. However, mapping of HIV testing uptake to granular geographic units was noted as a major gap and would be useful to inform granular targeting of HIV interventions. Furthermore, describing geographic patterns (through mapping) of the reasons for low testing uptake, in order to inform tailored HIV testing strategies, was also noted as a major gap. This study, therefore, uniquely mapped HIV testing uptake at granular geographic units (sub-location level), and further mapped geographic patterns of the reasons for low testing uptake (i.e., clients not found at home and declining testing).

To address the gaps noted, this study evaluated the following strategies: a HIV predictive risk score screening algorithm for use in the outpatient department that is context specific to the sub-Sahara Africa setting; geospatial analysis of new HIV diagnoses to granular geographic units; and granular mapping of HIV testing uptake and geographic patterns of the reasons for low testing uptake.

## 1.3 Study objectives

### 1.3.1 General objective

To evaluate a HIV predictive algorithm, geospatial analysis of new HIV diagnoses and mapping of HIV testing uptake in Homa Bay, Kisumu, and Siaya Counties, western Kenya.

### 1.3.2 Specific objectives

i. To evaluate the use of a HIV predictive risk-score screening algorithm identifying sub-populations with higher HIV positive yield to inform targeting of HIV testing among persons $\geq$15 years in Homa Bay, Siaya, and Kisumu Counties.

ii. To evaluate the use of geospatial analysis of new HIV diagnoses in identifying granular-geographic areas with higher HIV positive yield to inform targeting of HIV testing among persons ≥15 years in Siaya County.

iii. To evaluate the use of mapping HIV testing uptake in identifying granular-geographic areas with low HIV testing uptake yet higher HIV positive yield to inform targeting of HIV testing among persons ≥15 years in Siaya County.

### 1.3.3 Specific research questions

i. What is the use of a HIV predictive risk-score screening algorithm in identifying sub-populations with higher HIV positive yield to inform targeting of HIV testing among persons ≥15 years in Homa Bay, Siaya, and Kisumu Counties?

ii. What is the use of geospatial analysis of new HIV diagnoses in identifying granular-geographic areas with higher HIV positive yield to inform targeting of HIV testing among persons ≥15 years in Siaya County?

iii. What is the use of mapping of HIV testing uptake in identifying granular-geographic areas with low HIV testing uptake yet higher HIV positive yield to inform targeting of HIV testing among persons ≥15 years Siaya County?

## 1.4 Significance of the study

HIV testing is the cornerstone for identifying HIV positive individuals to link them to ART. Although globally and in sub-Sahara Africa, progress has been made in increasing ART coverage, many people are still not on ART. Those not on ART are a major contributor to HIV-related morbidity and mortality, and continued HIV transmission. By 2019,about 12.6million and 7.6 million people were not on ART globally and in sub-Sahara Africa, respectively (Joint United Nations Programme on HIV/AIDS, 2020b). Similarly, Kenya had an ART gap of about

310,000 people living with HIV not on ART, and in the three counties of Homa Bay, Kisumu and Siaya, about 53,000 people (President's Emergency Plan for AIDS Relief, 2020). Closing these gaps is critical in order to achieve HIV epidemic control.

As countries are making progress in increasing ART coverage, more people are initiating ART, and increasingly fewer people living with HIV in the population remain undiagnosed. Consequently, the HIV positivity (or yield) in HIV testing services is steadily declining in many high burden settings (World Health Organization, 2019). Furthermore, the 2015 WHO and Kenya HIV testing guidelines recommended programs to offer HIV testing in a universal, non-targeted manner. As more people initiate ART, and fewer people remain undiagnosed, continuing to offer HIV testing in a universal manner becomes inefficient, as many people are tested, hence programs use a lot of resources, and yet a low proportion of those who are HIV positive are identified. Strategies that help to target HIV testing to sub-populations or geographic areas that are more likely to yield a higher proportion of HIV positive individuals are critically needed, and have been highlighted as a key priority by the World Health Organization (Quinn C., 2020).

This study evaluated three strategies to tease out and identify sub-populations and granular geographic areas with higher HIV positive yield to inform the targeting of HIV testing for efficiency: a HIV predictive algorithm, geospatial analysis of new HIV diagnoses, and mapping of HIV testing uptake.

The study findings will inform the development of HIV testing screening tools that identify sub-populations with higher HIV-risk, to whom testing should be targeted. The results from this study’s assessment of a HIV predictive risk-score algorithm were included in a WHO webinar

held in June 2021, and a systematic review of HIV risk-based screening tools published in June 2021 (Ong et al., 2021; Quinn C., 2020). This study's results have therefore formed part of the evidence used to inform global policy on HIV screening tools (Ong et al., 2021). Following this, new HIV testing guidelines, released by WHO in July 2021 (World Health Organization, 2021), and Kenya in December 2022 (Ministry of Health National AIDS & STI Control Program, 2022), took into consideration recommendations from wide-WHO consultations, that included and referenced my publication in addition to other publications. The newer WHO and Kenya guidelines recommend provision of efficient targeted HIV testing.

Geospatial analysis and mapping are useful in identifying granular-geographic areas with higher HIV positive yield and low testing uptake, where HIV testing should be targeted for efficiency. Programs currently do not routinely conduct geospatial or mapping analysis, despite having a large amount of routine program data that they could potentially use. The results from this study will inform the development and implementation of policies on integrating geospatial analysis and mapping into routine program data analysis and use. Implementation will require training of staff and program capacity building (including acquisition of analytic software and tools) to conduct geospatial analysis and mapping.

The results of this study have been published in peer-review journals (Muttai, Guyah, Achia, et al., 2021; Muttai, Guyah, Musingila, et al., 2021), and will be presented at HIV programmatic meetings in Kenya (at county and national levels), and disseminated to HIV Implementing Partners, the President's Emergency Plan for AIDS Relief, and the Kenya Ministry of Health. The results will be used to inform global, national, and county government policies and strategies for targeting HIV testing, in order to ensure efficient use of resources and maximal epidemiologic impact.

### 1.5 Conceptual Framework

This is an implementation science study. Implementation science is the scientific study of methods to promote the systematic uptake of research findings or other evidence-based practice into routine health care to improve the quality and effectiveness of health services (Eccles & Mittman, 2006). There are three overarching aims to the use of implementation science that have been delineated: to describe and/or guide the process of translating research into practice; to understand and/or explain what influences implementation outcomes; and to evaluate implementation (Nilsen, 2020). From these three aims, five categories of theoretical approaches used in implementation science have been proposed: process models; determinant frameworks; classic theories; implementation theories; and evaluation frameworks (Nilsen, 2020). Process models describe and/or guide the process needed for translating research into practice (Nilsen, 2020). Determinant frameworks describe general types of determinants that are hypothesized or have been found to influence implementation outcomes; with each type of determinant typically comprising a number of individual barriers (hinders, impediments) and/or enablers (facilitators), which are seen as independent variables that have an impact on implementation outcomes i.e. the dependent variable (Nilsen, 2020). Classical theories are those that apply theories from fields outside of implementation science, such as psychology, sociology and organizational theory, and are applied to provide a better understanding and explanation of aspects of implementation (Nilsen, 2020). Implementation theories were developed by implementation researchers to provide a better understanding and explanation of aspects of implementation (Nilsen, 2020). Evaluation frameworks provide a structure to evaluate aspects of implementation science (Nilsen, 2020).

This study falls under the "determinant frameworks" theory, as it describes determinants (**independent variables**), that when applied to facilitators (in this study **mediating variables**), influence the impact on implementation outcomes (**dependent variable**). To fit into Nilsen's description of the aims of implementation science (Nilsen, 2020), overall, the aim of this study was to assess strategies that influence the implementation outcome. For this study, the strategies evaluated were: a HIV predictive risk-score algorithm; geospatial analysis of new HIV diagnoses; and mapping of HIV testing uptake (Figure 1.1).

The three strategies that were evaluated in this study, were derived from multiple **independent variables**: demographic factors, socioeconomic factors, behavioral factors, HIV testing-related factors and spatial factors. To explain this in more detail, the HIV predictive risk-score algorithm was developed from socio-demographic factors (age, occupation, marital status), sexual/behavioral factors (number of sexual partners, change in sexual partners), and HIV testing and related factors (time when last tested, and presence of tuberculosis, STI and recent HIV exposure). Geospatial analysis used spatial data (geo coordinates and geographic administrative unit shape-files), socio-demographic factors (age, sex, marital status), and HIV testing related factors (time since last HIV test). Mapping of HIV testing uptake used spatial data (geo coordinates and geographic administrative unit shape-files) and socio-demographic factors (age, sex).

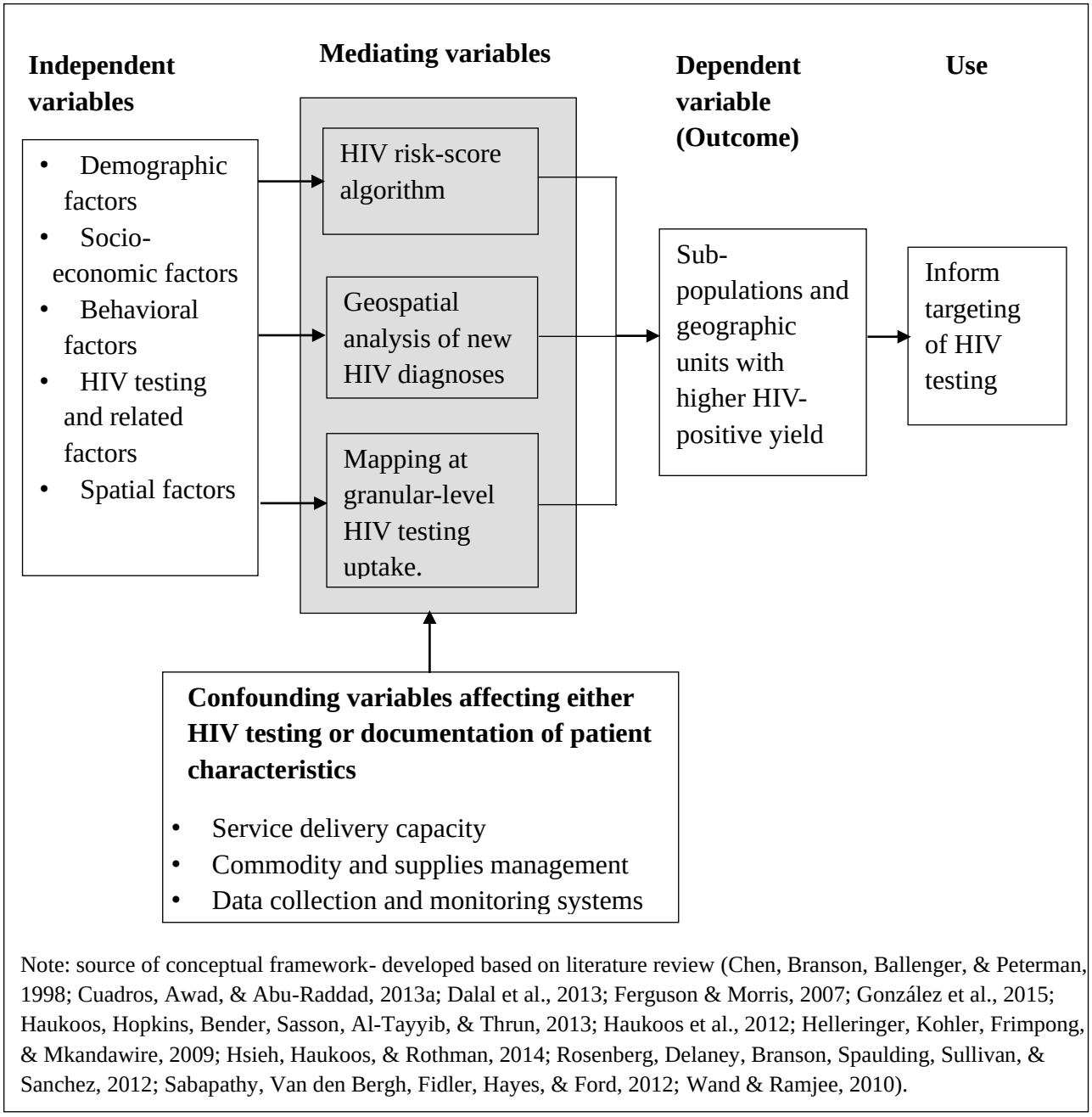
The three strategies, in turn were applied and influenced the identification of sub-populations and geographic units with higher HIV positive yield. In this role, the three strategies were **mediating variables**, as they explain and impact the process through which the independent variables affect the dependent variable. The rationale for using the mediating variables has been described in the "background" and "statement of the problem" sections of this thesis. The mediating variables

were derived from literature review, showing the value of HIV testing eligibility screening algorithms in identifying individuals with higher HIV-risk to be offered HIV testing (Chen et al., 1998; Haukoos et al., 2013; Haukoos et al., 2015; Haukoos et al., 2012; Hsieh, Haukoos, & Rothman, 2014; Rosenberg et al., 2012); the value of disease geospatial analysis and mapping in directing geographic regions to focus delivery of services (Cuadros, Awad, & Abu-Raddad, 2013a; Ferguson & Morris, 2007; González et al., 2015; Wand & Ramjee, 2010); and the value of geographic mapping using routine program data to inform interventions to improve HIV testing uptake (Alem, Liyew, & Guadie, 2021; Bassett et al., 2015; Nutor, Duah, Duodu, Agbadi, Alhassan, & Darkwah, 2021).

The **dependent variable** or outcome was identification of sub-populations and geographic units with higher HIV positive yield. Applying the three mediating variables influenced the dependent variable; and were useful to inform targeting of HIV testing for efficiency.

Multiple **confounding factors** that would potentially either impact HIV testing uptake, or lead to a low proportion of client characteristics being documented, were considered. The confounding factors considered were the capacity of health facilities and community structures to provide HIV testing (including human resource and technical skills), management and availability of commodities and supplies (mainly HIV test kits and tools for recording and reporting), and data collection and monitoring systems. Sub-optimal testing uptake or low documentation of client characteristics would have a major impact on the mediating variables. For example, this study used data for patients tested for HIV at six health facilities to develop and validate a HIV predictive algorithm; and if the testing coverage in these facilities were low, this potentially would have led to a bias in the study analysis, as it's likely clients with certain characteristics (e.g., those with clinical symptoms suggestive HIV infection) may be the ones prioritized and

offered testing. Additionally, in home-based HIV testing, if the testing coverage was low, or many sub-locations were not covered by home-based testing, potentially it would not have been possible to conduct geospatial analysis describing a whole population's geographic pattern of new HIV diagnoses. Confounding variables were therefore controlled in this study. For the six study sites used for HIV predictive algorithm development, confounding factors were controlled as follows: a) the health facilities chosen for the study were those with high testing coverage; b) Bondo County Hospital in Siaya was initially considered for inclusion in the study, but was later found to inconsistently document behavioral risk information and was therefore excluded; and c) at the six health facilities, data for an entire month were excluded if ≥50% of patients tested for HIV in that month did not have any documentation of behavioral risk characteristics. For the home-based testing analysis in Siaya County, confounding factors were controlled as follows: 18 sub-locations where <50% of households were enumerated for home-based testing were excluded from the analysis.




Note: source of conceptual framework- developed based on literature review (Chen, Branson, Ballenger, & Peterman, 1998; Cuadros, Awad, & Abu-Raddad, 2013a; Dalal et al., 2013; Ferguson & Morris, 2007; González et al., 2015; Haukoos, Hopkins, Bender, Sasson, Al-Tayyib, & Thrun, 2013; Haukoos et al., 2012; Helleringer, Kohler, Frimpong, & Mkandawire, 2009; Hsieh, Haukoos, & Rothman, 2014; Rosenberg, Delaney, Branson, Spaulding, Sullivan, & Sanchez, 2012; Sabapathy, Van den Bergh, Fidler, Hayes, & Ford, 2012; Wand & Ramjee, 2010).

**Figure 1.1: Conceptual framework to evaluate a HIV predictive algorithm, geospatial analysis of new HIV diagnoses and mapping of HIV testing uptake in Homa Bay, Kisumu, and Siaya Counties in Kenya**

### 1.6 Measurement of study variables

The variables that were measured in this study included independent, mediating, and dependent variables, as described in the conceptual framework. The independent variables included demographic factors, socioeconomic factors, behavioral factors, HIV testing-related factors and spatial factors. The mediating variables were a HIV predictive risk-score algorithm; geospatial analysis of new HIV diagnoses; and mapping of HIV testing uptake. The dependent variable or outcome was identification of sub-populations and geographic units with higher HIV positive yield. Table 1.1 shows the study variables, the measurement categories, and data sources.

In the measurement of variables in this study, steps were taken to minimize errors. Variable measurement error refers to the discrepancy between the true value of a variable and the measured value. In this study, measurement error could result from various reasons, such as human error, social desirability, poor data quality techniques, or incomplete data. In this study, measurement error was minimized through the following ways:

i. This study used data collected during the provision of routine HIV testing services in a retrospective cohort study design. Study personnel who handled the data were trained in data transcription, de-identification, and data transfer procedures that was standardized to minimize human errors.
ii. During provision of HIV testing services, to minimize social desirability bias, socio-demographic, behavioral and HIV testing-related information was obtained in a private location, and by trained counselors.
iii. Additionally, data collection during the provision of HIV testing services included data quality checks. During the analysis, data quality was also assessed.

iv. During univariable and multivariable regression analysis of the HIV predictive risk-score algorithm, missing data were omitted.

v. Reproducibility and validity of the HIV predictive risk-score algorithm were assessed using internal and external validation.

**Table 1.1: Measurement variables in study to evaluate a HIV predictive algorithm, geospatial analysis of new HIV diagnoses and mapping of HIV testing uptake in Homa Bay, Kisumu, and Siaya Counties**

| Variable | Measurement categories | Data source |
|---|---|---|
| **Independent variables for study objective 1** | | |
| **Socio-demographic characteristics** | | |
| Age | Continuous numeric | HIV Behavioral Questionnaire Implemented in HIV Testing (Appendix 1) |
| Sex | Male; Female | |
| Marital status | Never married; Married monogamous; Married polygamous; Cohabiting; Separated/divorced; Widowed | |
| Occupation | Professional/administrative/clerical; Manual (skilled and unskilled)/domestic; Agriculture; Trade/sales/service; Unemployed; School/college going | |
| **Behavioral characteristics** | | |
| Had sex in the prior 12 months | Yes/No | |
| Number of sexual partners in the prior 12 months | 1; $\geq$2 | |
| Changes in sexual partners in the prior 12 months | Widowed; Divorced/separated; Ended a sexual relationship; Newly marriage; New sexual partner; Not had change in sexual partner | |
| Had sex in exchange of money/favors in the prior 12 months | Yes/No | |
| Had sex under influence of alcohol/other substance in the prior 12 months | Yes/No | |
| Coerced to have sex in the prior 12 months | Yes/No | |
| Treated for a sexually transmitted infection (STI) in the prior 12 months | Yes/No | |
| Engaged in sex work, men who have with men, female anal sex, injecting drugs for pleasure in the prior 12 months | Yes/No | |

| Variable | Measurement categories | Data source |
|---|---|---|
| **HIV testing information** | | |
| Reason for HIV testing eligibility | Never been tested for HIV; HIV negative test >12 months prior; HIV negative test >6 to 12 months prior; HIV negative test 3 to 6 months prior; HIV negative test <3 months ago (unverified; HIV negative test date unknown; Has tuberculosis; STI or recent HIV exposure | Ministry of Health HIV testing register (Appendix 2) |
| **Independent variables for study objectives 2 and 3** | | |
| Enumerated client's relationship to household head | Household head; Spouse; Children ≥15 years; Other relatives; Non-relatives | Home-based HIV testing enumeration form (Appendix 4) |
| Sex | Male; Female | Ministry of Health HIV testing register (Appendix 5) |
| Age | Continuous numeric | |
| Marital status | Single; Married Monogamous; Married Polygamous; Separated/Divorced; Widow/Widower | |
| When last tested for HIV | <3 months; 3 - 12 months; >12 months; Never tested for HIV | |
| Spatial factors (geocodes and shape files) | | DIVA-GIS (https://www.diva-gis.org/gdata) |
| **Study mediating variables** | | |
| HIV predictive risk-score algorithm | | |
| Geospatial analysis of new HIV diagnoses | | |
| Mapping of HIV testing uptake | | |
| **Study dependent variable** | | |
| Identification of sub-populations and geographic units with higher HIV positive yield | | |

# CHAPTER TWO
# LITERATURE REVIEW

## 2.1 Need for efficient HIV testing strategies

Globally, there has been remarkable progress in increasing ART coverage. By 2019, 25.4 million of the 38.0 million people living with HIV were on ART, the number of people on ART tripling since 2010 (Joint United Nations Programme on HIV/AIDS, 2020c). Despite this progress, the 2020 global UNAIDS 90-90-90 targets were not met; since by the end of 2020, 84% of people living with HIV knew their HIV status, and about 73% were on antiretroviral therapy (Joint United Nations Programme on HIV/AIDS, 2021). Therefore, by 2020, globally 27% of people living with HIV were not on ART.

HIV testing services is the entry point to identifying individuals who are HIV positive to link them to ART. As countries are closing the ART gap, HIV testing becomes more challenging, since the HIV positive yield continues to decline, requiring programs to test many clients to identify few who are HIV positive. To ensure testing efficiency, strategies to help target HIV testing to sub-populations or geographic units that are more likely to yield a higher proportion of HIV positive individuals are an urgent global priority.

This study evaluated three strategies to identify sub-populations and granular-geographic areas with higher HIV positive yield, to inform targeting of HIV testing among persons $\geq$15 years of age, for program efficiency: a HIV predictive risk-score screening algorithm; use of geospatial analysis of new HIV diagnoses; and mapping of HIV testing uptake.

## 2.2 HIV testing policy recommendations

A concentrated HIV epidemic is where HIV has spread rapidly in a defined sub-population (such as men who have sex with men, sex workers, transgender people, people who use drugs or

people in prison or other closed settings) but is not well established in the general population (World Health Organization, 2015). On the other hand, a generalized HIV epidemic is when HIV is firmly established in the general population; and although sub-populations at high risk may contribute disproportionately to the spread of HIV, sexual networking in the general population is sufficient to sustain the epidemic.

The 2015 WHO HIV testing guidelines recommended that, in generalized epidemic settings, routine HIV testing should be offered to all clients (adults, adolescents, and children) in all clinical settings; and community testing offered, prioritized to key and priority populations. On the other hand, in low-level or concentrated epidemic settings, HIV testing should be offered to clients (adults, adolescents and children) in clinical settings who present with symptoms or medical conditions that could indicate HIV infection, including presumed and confirmed TB cases; and community testing offered to key populations. Additionally, regardless of epidemic type, routine HIV testing should be considered for malnutrition clinics, STI, viral hepatitis and TB services, and ANC settings, and for health services for key populations (World Health Organization, 2015).

Kenya has a generalized epidemic. Based on the WHO recommendations for generalized HIV epidemics (World Health Organization, 2015), the 2015 Kenya HIV testing guidelines (Kenya Ministry of Health, 2015) recommended that HIV testing and counseling should be offered to all patients attending health facilities, whether or not the patient has symptoms of HIV, and regardless of the reason for attending the health facility. Routine opt-out (offering HIV testing routinely to all clients as a standard of care, unless they decline) PITC should therefore be offered in all facilities and prioritized in all the service delivery points, including inpatient adult and pediatric facilities; outpatient facilities or departments; tuberculosis, sexually transmitted

infections, circumcision, post-rape care and family planning clinics; and maternal and child health clinics. In addition, the guidelines recommend provision of a strategic mix of community-based testing services.

### 2.3 Strategies evaluated in this study, to inform efficient HIV testing

This study evaluated three strategies used to tease out and identify sub-populations and granular-geographic areas that have higher HIV positive yield, to inform targeting of HIV testing among persons ≥15 years of age, for testing efficiency: a HIV predictive risk-score screening algorithm; geospatial analysis of new HIV diagnoses; and granular-level mapping of HIV testing uptake. These three strategies as discussed in each of the sub-sections below.

#### 2.3.1 Screening algorithms evaluated to predict HIV infection

Many studies have evaluated screening algorithms used in HIV testing services to identify individuals more likely to have HIV infection, who are then prioritized to be offered testing. These have included the use of algorithms among children, adolescents, key populations/other high-risk groups, women, and the general population.

***Screening algorithms to predict HIV infection among children***

A recent met analysis published in 2020 (Clemens, Macneal, Alons, & Cohn, 2020), conducted a systematic review and meta-analysis of studies evaluating symptom screening to identify children (0-15 years) eligible for further HIV testing in generalized epidemics. Several studies were identified, all were prospective or cross-sectional studies that developed and/or validated a screening tool to identify children at higher risk for being HIV infected. The studies, all conducted in settings of generalized epidemics, are described below.

A prospective cross-sectional study at Port Moresby General Hospital in Papua New Guinea developed a clinical algorithm to identify pediatric patients who should be offered HIV testing in

a setting of moderate HIV prevalence and limited resources (Allison, Kiromat, Vince, Handan, Graham, & Kaldor, 2011). The study identified independent predictors of HIV infection that included: persistent fever, lymphadenopathy, oral candidiasis and being underweight for age. The presence of any one of these conditions had a sensitivity of 96% in detecting a child with HIV infection. Using an algorithm based on the presence of at least one of these conditions resulted in around 40% of hospitalized children being offered testing. The clinical algorithm was therefore a useful screening tool for HIV infection in hospitalized children, and hence was recommended for use in situations where it's not feasible to offer universal HIV testing.

Another study done in Malawi (Moucheraud, Chasweka, Nyirenda, Schooley, Dovel, & Hoffman, 2018) assessed the sensitivity and specificity of a brief screening tool to identify at-risk children (aged 1–15 years) in the inpatient pediatric wards at 12 hospitals. The tool included the following parameters: ever been admitted to the hospital, had recurring skin problems, one or both natural parents had died, sicker more often than other children in the last 3 months, have frequent ear discharge, and shorter or smaller than others in the same age group. Frequent sickness was the most sensitive predictor of HIV status (55.1%), and having a deceased parent was the most specific (96.7%). False classification of HIV-negative status was rare, but occurred more often among boys and younger children.

In Zimbabwe a study was done to validated the performance of a screening tool in children aged 6-15 years attending primary healthcare facilities (Bandasona et al., 2016). The study reported a HIV prevalence of4.7%, and increased from 1.4% among those scoring zero on the tool to 63.6% among those scoring four. Using a score of not less than one as the cut-off for HIV testing, the tool had a sensitivity of 80.4%, a specificity of 66.3%, a positive predictive value of 10.4%, and

a negative predictive value of 98.6%. The number needed to screen to identify one child living with HIV would drop from 22 to 10 if this screening tool was used. The screening tool was a simple and sensitive method to identify children living with HIV in this setting.

A prospective study carried out on a cohort of 400 new patients attending the pediatric outpatient department in Medical College, Kolkata (Bandyopadhyay, Bhattacharyya, & Banerjee, 2009) assessed the feasibility of using a standardized questionnaire as a screening tool for detection of pediatric HIV at first contact. After examining, the attending physician noted his clinical impression, filled out the standardized questionnaire, and scored each patient, after which a HIV test was performed. Taking a score of 9 as the cut-off, the sensitivity and specificity of the scoring system was 95.7% and 98.6% respectively, showing that a clinic-epidemiological scoring system may be used to screen children for HIV in resource-limited settings.

In South Africa, a study was done to determine the validity of an algorithm used by primary care health workers; the HIV algorithm was implemented as part of the Integrated Management of Childhood Illness (IMCI), a strategy that aimed to improve childhood morbidity and mortality by improving care at the primary care level (C Horwood, Liebeschuetz, Blaauw, Cassol, & Qazi, 2003). The validity of the algorithm in detecting symptomatic HIV was compared with clinical diagnosis by a pediatrician and the result of a HIV test. Detailed clinical data were used to improve the algorithm. The pediatrician correctly identified 71.7% of children infected with HIV, whereas the IMCI/HIV algorithm identified 56.1%. Odds ratios (OR) were calculated to identify predictors of HIV infection and used to develop an improved HIV algorithm that was 67.2% sensitive and 81.5% specific in clinically detecting HIV infection. The study showed that

children with symptomatic HIV infection may be identified by primary level health workers using an algorithm.

Another similar study done in South Africa (C. Horwood, Vermaak, Rollins, Haskins, Nkosi, & Qazi, 2009) assessed the validity of an IMCI/HIV algorithm used by IMCI experts, the use of IMCI/HIV guidelines by IMCI trained health workers in routine clinical practice, and the burden of HIV among children under 5 years attending first level health facilities. The study found that IMCI experts using the HIV algorithm classified 71.1% HIV positive children as suspected symptomatic HIV, and 68.2% of the remaining children were identified as HIV exposed. The study findings showed that the HIV algorithm evaluated was a valid tool for identifying HIV infected and exposed children when correctly and comprehensively implemented.

***Screening algorithms to predict HIV infection among adolescents***

In Harare, Zimbabwe a study assessed an algorithm used by primary-care health workers to identify HIV-infected adolescents (10-18 years) in populations at high HIV-risk through mother-to-child transmission (Ferrand et al., 2011). The study found that HIV infection was independently associated with client-reported orphanhood, past hospitalization, skin problems, presenting with sexually transmitted infection and poor functional ability. Classifying adolescents as requiring HIV testing if they reported >1 of these five criteria had 74% sensitivity and 80% specificity for HIV, with the algorithm correctly predicting the HIV status of 79% of participants. In low-HIV-prevalence settings (<2%), the algorithm had a high negative predictive value (99.5%) and resulted in an estimated 60% decrease in the number of people needing to test to identify one HIV-infected individual, compared with universal testing.

Another study in Harare, that used data from a community-based HIV prevalence survey (Bandason et al., 2018), validated a 4-item (previous hospitalization, orphanhood, poor health status, and recurring skin problems) screening tool to identify adolescents aged 8-17 years living with HIV in healthcare facility settings. The 4-item screening tool had an area under the receiver operating curve of 0.65at a cut-off score ≥1. Its sensitivity was 56.3% and specificity 75.1%, positive predictive value was 2.9% and negative predictive value 99.2%. The number needed to test to diagnose one child using the screening tool was 55% lower than universal testing for HIV. Overall, the tool performed poorly.

Secondary analysis of data from a cohort of adolescent girls who were enrolled in the randomized control trial- HIV Prevention Trials Network (HPTN) 068 in rural South Africa, evaluated the utility of a risk score in predicting HIV incidence (Giovenco et al., 2019). The risk score was derived from the VOICE trial (Giovenco et al., 2019), and included the following variables: age, living with a primary partner, having a partner provide financial or material support, having a partner who has other partners, any alcohol use in the past three months, and herpes simplex virus type 2 serostatus. Scores ≥5 identified 85% of incident infections from 94% of the sample, compared to the VOICE sample in which scores ≥5 identified 91% of incident infections from only 64% of participants. The risk score did not predict HIV incidence after one year of follow-up (hazard ratio = 1.029) and showed poor predictive ability (area under the curve = 0.55). The study therefore shows that certain individual risk factors that comprise the risk score may be context specific or not relevant for adolescent populations.

***Screening algorithms to predict HIV infection among key populations***

Key populations comprise female sex workers, men to have sex with men, people who inject drugs and transgender populations. Many studies have assessed the use of HIV predictive

algorithms among key populations; most of the studies have been done among the MSM populations.

A study in the United States developed an evidence-based HIV risk assessment tool for MSM (Scott et al., 2020), using a large cohort of MSM, that included Black MSM. The final model included age (< 35, ≥ 35); Black race and Latino ethnicity; numbers of HIV-negative anal sex partners; number of insertive or receptive anal intercourse episodes; having one HIV-negative partner only; self-reported substance use; and bacterial sexually transmitted infection diagnosis. The model showed good discrimination in internal validation (C-statistic = 79.5). The external validation cohorts also showed good discrimination, with C-statistics ranging between 71.0 and 73.1.

Another study in the United states (in Atlanta Georgia), assessed the predictive ability of three published scores to predict HIV seroconversion in a cohort of black and white MSM (Jones, Hoenigl, Siegler, Sullivan, Little, & Rosenberg, 2017). The scores were the Menza score (Menza, Hughes, Celum, & Golden, 2009), the HIV Incidence Risk Index for MSM (Smith, Pals, Herbst, Shinde, & Carey, 2012) and the San Diego Early Test score (Hoenigl et al., 2015). The predictive ability of each score was low among all MSM and lower among black men compared to white men. Each score had lower sensitivity to predict seroconversion among black MSM compared to white MSM and low area under the curve values for the receiver operating characteristic curve indicating poor discriminatory ability. This study concluded that reliance on the available risk scores resulted in misclassification of high proportions of MSM, especially black MSM, in terms of HIV risk.

Among MSM in Beijing, China, a study created a composite score using questions from a routine survey to improve the estimation of HIV acquisition (Yin et al., 2018). The full penalized

model included 19 sexual predictors, while the reduced-form model had 12 predictors. The strongest predictors of HIV infection were non-Beijing residence, short-term living in Beijing, illegal drug use, multiple male sexual partners, receptive anal sex, inconsistent condom use, alcohol consumption before sex, and syphilis infection. Both models calibrated well; bootstrap-corrected c-indices were 0.70 (full model) and 0.71 (reduced-form model). This study therefore demonstrated that a validated risk score discriminated against higher-risk MSM.

In a multi-country HIV vaccine trial preparedness cohort study among individuals at high risk of HIV (Kansiime, Hansen, Hayes, & Ruzagira, 2023), identified factors were used to create and validate tools that predict HIV risk. The study comprised adults (18–45 years) at high-risk of HIV infection: female sex workers (FSW), and female and male fisher folk in Masaka, Uganda; female bar workers and FSW in Dar es Salaam and Mbeya, Tanzania; men who have sex with men (MSM), FSW and other at-risk individuals from the general population in Maputo, Mozambique; and the general population in areas of known high HIV incidence in Durban, South Africa. The tool had the following variables: age, sex, recreational drug use, unprotected male-to-male anal sex, a sexual partner who had other partners, transactional sex and having a partner who was a long-distance truck driver/miner. The HIV prediction tool created had good predictive ability [area under the curve (AUC) = 0.70, 95% CI 0.66–0.74].

On the Kenyan coast, a study used data from an open cohort, which followed 753 initially HIV-negative MSM participants for more than 1378.5 person-years, to develop an empiric risk score for targeting PrEP delivery (Wahome et al., 2018). Independent predictors of incident HIV infection in this cohort were the age of 18–24 years, having only male sex partners, having receptive anal intercourse, having any unprotected sex, and having group sex. A risk score of $\geq 1$ corresponded to an HIV incidence of $\geq 2.2$ and identified 81.3% of the cohort participants as

being at high risk for HIV acquisition. The area under the receiver operating characteristic curve was 0.76. The study found that this empiric risk score was useful in assessing HIV acquisition risk.

A study among MSM in the United States (Smith et al., 2012), developed and validated a rapid, risk screening tool for identifying persons at the highest risk of incident infection. The final logistic regression model included age, and the following behaviors reported during the past 6 months: total number of male sex partners, total number of HIV positive male sex partners, number of times the participant had unprotected receptive anal sex with a male partner of any HIV status, number of times the participant had insertive anal sex with an HIV positive male partner, whether the participant reported using poppers, and whether they reported using amphetamines. The area under the receiver operating characteristic curve was 0.74, possible scores on the index ranged from 0 to 47 and a score $\geq$10 had a sensitivity of 84% and a specificity of 45%, levels appropriate for a screening tool.

Another study done in the United States (Hoenigl et al., 2015) developed and validated a score to estimate incident HIV infection risk. Clinical and behavioral data collected within an acute and early HIV infection screening program were used to construct and validate a simple multivariable risk behavior score predictive of acute and earlier HIV infection among MSM. The San Diego Early Test score excluded demographics and focused instead on relevant current risk variables directly associated with HIV acquisition among MSM: condomless receptive anal intercourse, number of male partners within the previous 12 months, and bacterial STIs. Four risk behavior variables were significantly associated with an acute and early HIV infection diagnosis (i.e., incident infection) in multivariable analysis and were used to derive the San Diego Early Test score: condomless receptive anal intercourse with an HIV positive MSM (3

points), the combination of condomless receptive anal intercourse plus ≥5 male partners (3 points), ≥10 male partners (2 points), and diagnosis of bacterial sexually transmitted infection (2 points)—all as reported for the prior 12 months. The C-statistic for this risk score was >0.7 in both data sets, showing good performance.

***Screening algorithms to predict HIV infection among women***

To develop and validate an HIV risk assessment tool to predict HIV acquisition among African women (Balkus et al., 2016), data were analyzed from 3 randomized trials of biomedical HIV prevention interventions among African women (VOICE, HPTN 035, and FEM-PrEP) (Karim et al., 2011; Marrazzo et al., 2015; Van Damme et al., 2012). Standard methods for the development of clinical prediction rules were used to generate a risk-scoring tool to predict HIV acquisition over the course of 1 year. The final risk score resulting from multivariable modeling included age, married/living with a partner, partner providing financial or material support, partner having other partners, alcohol use, detection of a curable sexually transmitted infection, and herpes simplex virus 2 serostatus. Point values for each factor ranged from 0 to 2, with a maximum possible total score of 11. Scores ≥5 were associated with HIV incidence >5 per 100 person-years and identified 91% of incident HIV infections from among only 64% of women. The area under the curve (AUC) for the predictive ability of the score was 0.71, indicating good predictive ability. Risk score performance was generally similar to internal cross-validation (AUC = 0.69) and external validation (AUC range between 0.58 and 0.70).

Another study done among women in South Africa (Wand et al., 2018), reported 7-factors that were significant predictors of HIV infection: <25 years old, being single/not cohabiting, parity (<3), age at sexual debut (<16), 3+ sexual partners, using injectables and diagnosis with a

sexually transmitted infection. A score of C25 (out of 50) was the optimum cut point with 83% sensitivity in the development dataset and 80% in the validation dataset.

***Screening algorithms to predict HIV infection among the outpatient population***

The outpatient population are individuals who visit the outpatient department in a health facility or hospital to seek health services. The outpatient department is the part of a health facility or hospital designed for the treatment of people with health problems who visit the facility or hospital for diagnosis or treatment, but do not at the time require a bed or to be admitted for overnight care.

Although many patients globally flow through outpatient or emergency departments to seek health services, few screening algorithms have been evaluated for use in outpatient settings. A study done in the United States (Haukoos et al., 2012), derived and validated an instrument called "Denver Human Immunodeficiency Virus (HIV) Risk Score", to identify patients at risk for HIV infection, using patient data from a metropolitan sexually transmitted disease clinic in Denver, Colorado (1996–2008). Multivariable logistic regression was used to develop a risk score from 48 candidate variables using newly identified HIV infection as the outcome. Validation was performed using an independent population from an urban emergency department in Cincinnati, Ohio. The final score included age, gender, race/ethnicity, sex with a male, vaginal intercourse, receptive anal intercourse, injection drug use, and past HIV testing, and values ranged from -14 to +81. The risk score performed well: the calibration regression slope for the validation sample was 1.07 and $R^2$ was 0.98. The area under the receiver operating curve for the validation sample was 0.75.

Another study done in the United States (Lyons et al., 2013), compared targeted screening and universal testing, among patients attending the emergency department. Indications for targeting

contained over fifty items, including: 1) clinician identified signs and symptoms of HIV, and 2) clinician or counselor identified risk behaviors, homelessness, mental illness, STI exposure or infection, violence, substance use, pregnancy, and incarceration. The study concluded that targeted screening, even when fully implemented with maximally permissive selection, offered no important increase in positivity rate, or decrease in tests performed. Universal screening diagnosed more cases, because more were tested, despite a modestly lower consent rate.

A study done in Spain (Elías et al., 2016) developed a Spanish-structured HIV risk of exposure and indicator conditions (RE&IC) questionnaire. People attending an emergency room, or a primary clinical care center were offered to participate in a prospective, 1 arm, open label study, in which all enrolled patients filled out the developed questionnaire and were HIV tested. The questionnaire included risk exposure items (unprotected sexual intercourse, partner with HIV infection, man with man sex, have received any hem derivative transfusion, parental illicit or recreational drug use, any suspicion of HIV acquisition) and clinical conditions items (sexually transmitted infection, lymphoma, cancer, herpes zoster, mononucleosis-like syndrome, B or C hepatitis, thrombopenia, seborrheic dermatitis, candidiasis oral, oral hairy leukoplakia, unexplained fever, unexplained prolonged diarrhea (>3 months), unexplained weight loss, mycobacterium tuberculosis disease. HIV RE&IC questionnaire sensitivity was 100% to predict HIV infection, with a specificity of 49%. positive predictive value was 0.80%, and negative predictive value reached100%.

***A recent study in sub-Sahara Africa evaluating a screening algorithm to predict HIV infection among the outpatient population***

After the 2021 publication of this study's findings on a HIV predictive risk-score algorithm (Muttai, Guyah, Musingila, et al., 2021); results from one additional study were published in

2022. The study was conducted in Malawi (Moucheraud et al., 2022), and used exit survey data collected at outpatient departments to estimate the sensitivity, specificity, negative and positive predictive values of screening tools that include questions about sexual behavior and use of health services. The study compared a full tool (seven relevant questions), to a reduced tool (five questions, excluding sexual behavior measures), and to standard of care (two questions, never tested for HIV or tested > 12 months ago, or seeking care for suspected STI). Suspect STI and ≥ 3 sexual partners were associated with HIV positivity but had weak sensitivity and specificity. The full tool (using the optimal cutoff score of ≥ 3) achieved 55.6% sensitivity and 84.9% specificity for HIV positivity; the reduced tool (optimal cutoff score ≥ 2) achieved 59.3% sensitivity and 68.5% specificity; and standard of care 77.8% sensitivity and 47.8% specificity. This study concluded that screening tools for HIV testing in outpatient departments do not offer clear advantages over standard of care.

***Gap that this study's HIV predictive risk-score algorithm addresses***

As described above, many studies have evaluated HIV testing eligibility screening algorithms among children, adolescents, key populations, and women. Although many adult patients around the world flow through health facility outpatient departments, not many screening algorithms are available or have been evaluated for use in the facility-outpatient setting. Three studies conducted in the United States and Spain that evaluated algorithms for use among adults in the general outpatient setting, were reviewed. The "Denver HIV Risk Score" (Haukoos et al., 2012) was developed using patient data from a sexually transmitted disease clinic, and has been widely validated in the outpatient setting. Another study done in the United States (Lyons et al., 2013), compared targeted screening and universal testing, among patients attending the outpatient

emergency department. A study done in Spain (Elías et al., 2016) developed a Spanish-structured HIV risk of exposure and indicator conditions (RE&IC) questionnaire.

These three studies, although conducted in the outpatient setting, were all done in the settings of low HIV prevalence and concentrated HIV epidemics. Concentrated epidemics are where HIV has spread rapidly in a defined sub-population such as men who have sex with men, sex workers, transgender people, people who use drugs or people in prison or other closed settings, but is not well established in the general population (World Health Organization, 2015). The United States HIV epidemic is disproportionate; gay and bisexual men account for most new HIV infections (66%), and disparities in HIV remain severe among some racial and ethnic minority groups (Black or African American people face rates of infection that are eight times as high as White people; and Hispanic and Latino people face rates that are almost four times as high) (Centers for Disease Control and Prevention, 2023). Similarly, the HIV epidemic in Spain is driven by MSM who account for 64% of new infections, while heterosexual relations account for 31% of new infections (Rivero & Moreno, 2017). The risk factors for HIV in the United States and Spain therefore differ from those in the sub-Saharan Africa setting. Sub-Sahara Africa and Kenya have generalized epidemics. Generalized epidemics is where HIV is firmly established in the general population; and although sub-populations at high risk may contribute disproportionately to the spread of HIV, sexual networking in the general population is sufficient to sustain the epidemic.

HIV-risk factors in generalized epidemics are mostly related to heterosexual relationships in the general population, which differs from concentrated epidemics, where risk factors are mostly related to sexual relations within specific affected sub-populations. Developing a screening algorithm for use in the outpatient department, that is context specific to the sub-Saharan Africa region, that has a generalized epidemic, was noted as a major gap.

***Factors associated with HIV infection that were considered for inclusion in this study's development of a HIV screening predictive screening algorithm***

Heterosexual intercourse remains the main mode of HIV transmission in the sub-Saharan Africa setting (Kharsany & Karim, 2016). In the 2008 Kenya modes of transmission study (Kenya National AIDS Control Council, 2009), HIV transmissions were attributed to heterosexual sex within union (44%), casual heterosexual sex (20%), sexual workers (14%), men who have sex with men and prisons (15%), injecting drug use (3.8%) and health facility related (2.5%). In the former Nyanza region (that included Homa Bay, Siaya, Migori and Kisumu Counties that border Lake Victoria) additional transmission was attributed to the fishing communities.

Several studies have been conducted, describing the factors that are associated with HIV infection. A study conducted in Kenya (Kimani, Ettarh, Ziraba, & Yatich, 2013) showed that across both sexes, marital status was a significant risk factor for HIV infection; with married and formerly married individuals being less likely to have used condoms during the last sexual intercourse relative to never married individuals. The risk of contracting HIV in marriage has been associated with unsafe sexual practices, such as lack of condom use especially where partners are engaged in extramarital affairs. In another study (Kimani, Ettarh, Ziraba, & Yatich, 2011), married respondents (odds ratio 1.78; p value<0.05) and those who were divorced, separated, or widowed (odds ratio 4.06; p value<0.001) were significantly more likely to be infected with HIV compared to respondents who were never married. Men who were circumcised (odds ratio 0.36; p value<0.05) were less likely to be HIV positive compared to those who were not circumcised. A study by Amornkul et al. (Amornkul et al., 2009) showed that HIV infection was strongly associated with age, a higher number of sexual partners, widowhood, and herpes simplex virus type 2 seropositivity. Another study (Nalugoda et al., 2014), reported that having multiple sexual partners significantly increased the risk of HIV

acquisition in both women and men. Analysis of survey data in Kenya (Kimanga, Ogola, & Umuro, 2014), found that among women, factors associated with undiagnosed HIV infection included being aged 35–39 years, divorced or separated, from urban residences and Nyanza region, self-perceiving at moderate risk of HIV infection, condom use with the last partner in the previous 12 months, and reporting 4 or more lifetime number of partners; while among men, widowhood, condom use with the last partner in the previous 12 months, and lack of circumcision were associated with undiagnosed HIV infection. Oluoch et al. in a similar analysis of nationally representative survey data (Oluoch et al., 2011) reported the following factors as independently associated with HIV among women: region (Nyanza vs Nairobi), number of lifetime sexual partners (6-9 *vs* 0-1 partners), herpes simplex type 2 virus, marital status (widowed *vs* never married) and consistent condom use with last sexual partner. Among men, correlates of HIV infection were the 30-to-39 year-old age group, number of lifetime sexual partners (10+ *vs* 0-1 partners), herpes simplex type 2 virus, syphilis, consistent condom use with last sexual partner and lack of circumcision.

Intimate partner violence is defined as "behavior within an intimate relationship that causes physical, sexual, or psychological harm, including acts of physical aggression, sexual coercion, psychological abuse and controlling behaviors" (World Health Organization, 2010). Population-based data analysis from the 2005 Rwanda Demographic and Health Survey (Dude, 2011) indicated that women with few, if any, other sexual risk factors who have experienced sexual, physical, or emotional abuse within their marriages were 1.61–3.46 times as likely to test positive for HIV, and 2.14–4.11 times more likely to report another STI.

Sexually transmitted infections have been associated with increased HIV transmission (Amornkul et al., 2009; Pettifor et al., 2005). STIs that cause ulcers generally increase the

shedding of HIV in the genital tract (Coombs, Reichelderfer, & Landay, 2003; Fleming & Wasserheit, 1999; Mbopi-Kéou et al., 2000; Røttingen, Cameron, & Garnett, 2001). Studies on female sex workers showed a 3.9 times higher shedding of HIV in the presence of vaginal or cervical ulcers (Ghys et al., 1997). Genital ulcer disease can also affect HIV levels in semen by affecting systemic viral loads or increasing local inflammation. In a study in Malawi, men with genital ulcers and non-gonococcal urethritis were found to shed higher amounts of HIV in semen compared with men with urethritis alone (Dyer et al., 1998). Even asymptomatic urethritis has been associated with HIV shedding in semen (Winter et al., 1999). The effects of STIs on susceptibility to HIV are supported by many studies that link a history of an STI to HIV acquisition (Coombs, Reichelderfer, & Landay, 2003; Fleming & Wasserheit, 1999; Røttingen, Cameron, & Garnett, 2001).

Studies evaluating the protective effect of male circumcision reported an incidence rate of 0.85 per 100 person-years in male circumcised, and 2.1 per 100 person-years in those not circumcised; corresponding to a relative risk (RR) of 0.40 and protection of 60% (Auvert, Taljaard, Lagarde, Sobngwi-Tambekou, Sitta, & Puren, 2005). A similar study (Bailey et al., 2007), reported a HIV incidence of 2·1% in the circumcision group and 4·2% in the uncircumcised group, the relative risk of HIV infection in circumcised men was 0·47, corresponding to a reduction in the risk of acquiring an HIV infection of 53%. The reasons for the protective effect of circumcision on HIV acquisition have been reported, and several direct or indirect factors explain this (Szabo & Short, 2000). Direct factors may be keratinization of the glans when not protected by the foreskin, short drying after sexual contact, reducing the life expectancy of HIV on the penis after sexual contact with an HIV positive partner, reduction of the total surface of the skin of the penis, and reduction of target cells, which are numerous on the

foreskin (Patterson et al., 2002). Indirect factors may be a reduction in the acquisition of other STIs, which in turn will reduce the acquisition of HIV.

### 2.3.2 Geospatial analysis of new HIV diagnoses

Geospatial analysis is the manipulation of data based on geographic location. Several studies have described the use of geospatial analysis to describe HIV prevalence, incidence, transmission/infection, HIV treatment cascade, and testing uptake.

#### *Geospatial analysis of HIV prevalence*

A study assessing prevalent HIV cases in Atlanta (Hixson, Omer, Del Rio, & Frew, 2011), to examine case distribution trends and population characteristics at the census tract level that may be associated with clustering effects; identified one large cluster centralized in downtown Atlanta that contained 60% of prevalent HIV cases. The prevalence rate within the cluster was 1.34% compared to 0.32% outside the cluster. Clustered tracts were associated with higher levels of poverty (OR = 1.19), lower density of multi-racial residents (OR = 1.85), injection drug use (OR = 1.99), men having sex with men (OR = 3.01), and men having sex with men and intravenous drug use (OR = 1.6). This study therefore noted that the HIV epidemic in Atlanta is concentrated in one large cluster characterized by poverty, men who have sex with men (MSM), and intravenous drug usage.

In a high prevalence rural population in KwaZulu-Natal, South Africa, a study investigating the micro-geographical patterns and clustering of HIV infections (Tanser, Bärnighausen, Dobra, & Sartorius, 2017), used a two-dimensional Gaussian kernel of 3-kilometer radius to produce robust estimates of HIV prevalence that vary across continuous geographical space. The study found considerable geographical variation in local HIV prevalence (range = 6–36%) within this relatively homogenous population and provided clear empirical evidence for the localized

clustering of HIV infections. Three high-risk, overlapping spatial clusters [Relative Risk (RR) = 1.34–1.62] were identified by the Kulldorff statistic along the National Road ($P \leq 0.01$), whereas three low-risk clusters (RR = 0.2–0.38) were identified elsewhere in the study area ($P \leq 0.017$). The findings showed the existence of several localized HIV epidemics of varying intensity that are partly contained within geographically defined communities. Another study in South Africa, (Wand & Ramjee, 2010) assessed the core areas of HIV infection in KwaZulu-Natal, South Africa, using epidemiological data among sexually active women from localized communities. The study identified three hotspots with excessively high HIV prevalence rates of 56%, 51% and 39%; reinforcing the inference that the risk of HIV infection is associated with definable geographical areas.

In an effort to clarify specific drivers of HIV transmission and identify priority populations for HIV prevention interventions, a study was done to conduct comprehensive mapping of the spatial distribution of HIV infection across sub-Saharan Africa (Cuadros, Awad, & Abu-Raddad, 2013b). The study used data from Demographic and Health Surveys conducted in 20 countries, and used a maximum circular window of a 100 kilometers radius for scanning potential clusters with high or low numbers of HIV infections. The results showed stark geographic variations in HIV transmission patterns within and across countries of sub-Sahara Africa. About 14% of the population in sub-Sahara Africa is located in areas of intense HIV epidemics. Meanwhile, another 16% of the population is located in areas of low HIV prevalence, where some behavioral or biological protective factors appear to have slowed HIV transmission. Another study that included 7 countries in Eastern and Southern Africa: Kenya, Malawi, Mozambique, Tanzania, Uganda, Zambia, and Zimbabwe (Bulstra et al., 2020), mapped and characterized high-prevalence areas for young adults (15-29 years of age), as a proxy for areas with high levels of

transmission. The study found that, at the sub national level, there were areas with prevalence among young adults as high as 11% or 15% alternating with areas with prevalence between 0% and 2%, suggesting the existence of areas with high levels of transmission. Overall, 15.6% of heterogeneity could be explained by an interplay of known behavioral, socioeconomic, and environmental factors. Maps of the interpolated random effect estimates show that environmental variables, representing indicators of economic activity, were most powerful in explaining high-prevalence areas.

A study in Kenya (Waruru et al., 2018) aimed to identify geographic clusters with significantly higher HIV prevalence to focus interventions, using a defined sizeable scan window with a maximum diameter of 100 kilometers. The study found that about half of the survey locations, 112/238 (47%) had high rates of HIV (HP clusters), with 1.1–4.6 times greater adults living with HIV observed than expected. Richer persons, compared with respondents in the lowest wealth index, had higher odds of belonging to a HP cluster; respondents who perceived themselves to have greater HIV risk or were already HIV-infected had higher odds of belonging to a HP cluster compared with perceived low risk. Men who had ever been clients of female sex workers had higher odds of belonging to a HP cluster than those who had never been; and uncircumcised men vs circumcised. The study therefore found that HIV infection in Kenya exhibits localized geographic clustering associated with socio-demographic and behavioral factors, suggesting disproportionate exposure to higher HIV risk. The study further noted that most of the high-prevalence clusters were near a lake/river, major road/highway, an economic hub, or highly productive agricultural zones such as tea growing areas and flower farms. High prevalence clusters in the Nairobi region were in informal settlements. The study indicated that there are

pockets of higher HIV infection that otherwise may not be well described in a generalized and spatially diffused epidemic.

***Geospatial analysis of HIV incidence/ transmission***

In China (Zhu et al., 2021) a study aimed to develop quantitative analytic measures for accurately identifying hot-spot areas in the growth of new HIV infections. The study found that geographic location of HIV cases had an uneven distribution along major roads and clustered at road intersections. The geographic mapping showed that several areas were clustered with more recently infected HIV cases than long-term infected cases. The quantitative analyses found twenty-three townships showing an increase in the number of recent infections.

A study in rural South Africa (Cuadros et al., 2022) hypothesized that HIV geographical clusters (geospatial areas with significantly higher numbers of HIV positive individuals) can behave as highly connected nodes in the transmission network. Continuous surface maps of HIV prevalence and HIV seroconversion were generated using a moving two-dimensional Gaussian kernel of a 3-kilometer search radius to produce robust epidemiological estimates that vary across continuous geographical space. The study found that more than 70% of the HIV transmission links identified were directly connected to a HIV geographical cluster located in a peri-urban area. Moreover, the study identified a single central large community of highly connected nodes located within the HIV cluster. This nodule was composed of nodes highly connected among them, forming a central structure of the network that was also connected with the small sparser nodules located outside of the HIV geographical cluster. This study supported the evidence of the high level of connectivity between HIV geographical high-risk populations and the entire community. The effect of mobility and migration intensity in predicting HIV

acquisition risk in high-incidence communities near a major road was described in South Africa (Tanser et al., 2017).

***Geospatial analysis describing the HIV treatment cascade***

Spatial analysis in Philadelphia was useful in characterizing the HIV treatment cascade; the study specifically identified spatial patterns as a strong independent predictor of linkage to care, retention in care, and viral suppression (Eberhart et al., 2013).

***Geospatial analysis describing access to HIV services***

A study done in South Africa showed that the implementation of a HIV/AIDS Spatial Information Management System played a critical role in determining where and when to offer HIV services, improving the quality of care for HIV positive patients, increasing accessibility of services and delivering a cost-effective mode of information (Busgeeth., 2004).

Using data from the estimated number of adults living with HIV aged 15-49 years in 47 countries in sub-Saharan Africa and the global map of travel time to the nearest health care facility by motorized and non-motorized transportation, a study aimed to generate high-resolution maps of underserved areas where people cannot access the closest health care facilities within appropriate travel time in sub-Saharan Africa (H. Kim, Musuka, Mukandavire, Branscum, & Cuadros, 2021). The study identified and mapped more than 7 million people living with HIV in the areas that lacked access to health care within a 10-minute travel time, and 1.5 million people living with HIV in the areas that lacked access to health care within a 60-minute travel time. The identified locations of underserved areas were an indicator of the challenge faced by people living with HIV in accessing health services in sub-Sahara Africa.

***Geospatial analysis describing HIV testing uptake***

In South Africa, a study compared the yield, geographic distribution, and demographic characteristics of populations tested by mobile- and clinic-based HIV testing programs deployed by iThembalabantu Clinic in Durban (Bassett et al., 2015). The study found that the HIV prevalence at mobile sites ranged from 0% to 26%.

***Factors contributing to geographic variation in new HIV diagnoses***

A study in Kenya (Waruru et al., 2021a) analyzed facility-level HIV testing data to assess the spatial distribution of newly diagnosed HIV positive individuals, setting a circular scan window of 50-kilometer radius. The study found that most facilities (3,034, 76.4%) were not spatially autocorrelated for the number of newly diagnosed HIV positive individuals. The study identified 123 clusters with a significantly high number of newly diagnosed HIV-infected persons, of which 73 (59.3%) were not in the five highest HIV-burden counties in the country. Clusters with a high number of newly diagnosed persons had twice the number of positives per 1,000,000 tests than clusters with lower numbers (29,856 vs. 14,172).

***Gap that this study's geospatial analysis of new HIV diagnoses addresses***

Geospatial analysis and mapping of new HIV diagnoses to smaller geographic units, that would be useful to inform more granular targeting of HIV interventions, was noted as a major gap. This study explored geospatial analysis of new HIV diagnoses to the smallest possible geographic unit. This study further described clusters of higher and lower new HIV diagnoses, overlayed them with ecological features, and additionally described factors associated with higher HIV diagnoses in a spatially integrated Bayesian model.

### 2.3.3 Granular-level mapping of HIV testing uptake

***Home-based HIV testing uptake and associated factors***

A meta-analysis of studies published between 2000 and 2012 that described home-based HIV testing in sub-Saharan Africa (Sabapathy et al., 2012), assessed the proportion of individuals accepting testing. The studies came from five countries: Uganda, Malawi, Kenya, South Africa, and Zambia. The proportion of people who accepted home-based testing ranged from 58.1% to 99.8%.

Many studies have assessed the factors associated with home-based testing. In the metanalysis (Sabapathy et al., 2012), forty-eight percent of the individuals offered testing were men, and they were just as likely to accept home-based testing as women (pooled odds ratio=0.84; 95% CI: 0.56–1.26). A study done in Malawi measured the uptake of home-based testing among members of the poorest households (Helleringer et al., 2009), and found that members of households in the lowest income quartile were significantly less likely to have ever used facility-based testing services than the rest of the population (odds ratio = 0.60, 95% confidence interval (CI): 0.36 to 0.97). In contrast, they were significantly more likely to use home-based testing services provided during the study (adjusted odds ratio = 1.70, 95% CI: 1.04 to 2.79). The differences in the uptake of home-based testing were not due to underlying differences in socioeconomic characteristics or HIV risk factors. A study in Uganda investigated the level of acceptance of home-based HIV testing and the factors associated with acceptance in an urban setting (Sekandi et al., 2011). The study reported a HIV testing uptake of 66%, and found that being male (odds ratio 1.65; 95% CI 1.03, 2.73), age 25-34 (adjusted odds ratio 0.63; 95% CI 0.40, 0.94) and ≥35 years (adjusted odds ratio 0.30; 95% CI 0.17, 0.56), being previously married (adjusted odds ratio 3.22; 95% CI 1.49, 6.98) and previous HIV testing (adjusted odds ratio 0.50; 95% CI 0.30, 0.74) were significantly associated with home-based testing acceptance. The study further

reported reasons for not accepting home-based testing as not being emotionally prepared and having to consult spouses or parents. A South African study (Naik, Tabana, Doherty, Zembe, & Jackson, 2012b) reported 75% testing uptake, and the reasons for not testing included not being ready/feeling scared/needing to think about it (34.1%); and not feeling at risk of having or acquiring HIV (10.1%). A study that aimed to determine factors contributing to the acceptability of home-based HIV counseling and testing among commuters in Johannesburg's inner city (Muloongo, Tshuma, Chimoyi, Setswe, Sarfo, & Nyasulu, 2014), reported home-based testing acceptability of 64%. High school education (adjusted odds ratio 0.61, CI: 0.46-0.85), inner city residence (adjusted odds ratio 0.70, CI: 0.52-0.94), previous HIV testing in the hospital (adjusted odds ratio 0.22, CI: 0.15-0.32) and at home (adjusted odds ratio 0.18, CI: 0.11-0.27) were significantly less likely associated with home-based testing acceptability. On the other hand, being married (adjusted odds ratio 1.64, CI: 1.15-2.32), recent HIV testing (adjusted odds ratio 1.85, CI: 1.15-2.99) and having experienced negative health worker attitude (adjusted odds ratio 2.41, CI: 1.66-3.48) were significantly more likely associated with home-based testing acceptability. A Ugandan study (Ruzagira, Baisley, Kamali, & Grosskurth, 2018) reported high home-based testing uptake of 89.9%; uptake being higher among men (adjusted odds ratio 1.20, 95% CI 1.07-1.36) than women, and decreasing with increasing age. Although more women than men were found at home in the study, men were more likely than women to accept home-based testing.

A possible reason for this finding was that compared to men, women are more likely to attend health care facilities and consequently to learn about their HIV status through facility-based testing programs (Sekandi et al., 2011). Other studies have also reported home-based testing uptake being the highest in the youngest age group (Dalal et al., 2013; Helleringer, Kohler,

Frimpong, & Mkandawire, 2009). Possible reasons for this include a lower HIV risk perception in older compared to younger persons (Sekandi et al., 2011). Client-level reasons for declining home-based HIV testing have been reported to include: not being ready/feeling scared/needing to think about it, fear of knowing one's status, not feeling at risk of having or acquiring HIV, preferring to test away from home, and wanting to test later (Dalal et al., 2013; Kranzer et al., 2008; Naik, Tabana, Doherty, Zembe, & Jackson, 2012a).

***Spatial analysis and mapping of home-based HIV testing uptake***

A study done in Kenya reported near similar home-based testing uptake between two geographic regions- a rural (Lwak) and urban informal settlement (Kibera) as 82.8% and 80.8% respectively (Dalal et al., 2013). A study in South Africa (Bassett et al., 2015) compared the yield, geographic distribution, and demographic characteristics of populations tested by mobile- and clinic-based HIV testing programs deployed by iThembalabantu Clinic in Durban, South Africa. The study found that the HIV prevalence at mobile sites ranged from 0% to 26%, and clinic-based testers traveled further than the clinic closest to their home to test; and mobile-based testers were more likely to test ≥ 5 kilometers away from home. A study done in Ghana (Nutor et al., 2021) developed an HIV testing prevalence surface map using spatial interpolation techniques to identify geographical areas with low and high HIV testing. The surface map revealed intra-regional level differences in HIV testing estimates. In Ethiopia (Alem, Liyew, & Guadie, 2021) a study found that the spatial patterns of home-based testing uptake were non-random (Global Moran's $I = 0.074$, $p$ value$< 0.001$). Forty-seven primary clusters were identified that were in the entire Somali region with a relative likelihood of 1.50 and a Log-Likelihood Ratio of 135.57. Using the 2003 Nigerian National Demographic and Health Survey (Nwachukwu & Odimegwu, 2011), the regional prevalence, pattern and correlates of voluntary counseling and testing for

HIV among youths aged 15 to 24 years in Nigeria were examined. Results showed that the national prevalence of voluntary testing was low (2.6%) with regional variations. Generally, the critical factors associated with voluntary testing uptake were age, sex, education, wealth index and risk perception with North (sex, education, religion, occupation, and risk perception) and South regional (age and education) variations.

***Gap that this study's mapping of home-based HIV testing uptake addresses***

Mapping of HIV testing uptake to granular geographic units was noted as a major gap and would be useful to inform granular targeting of HIV interventions. Furthermore, geographically mapping the reasons for low testing uptake, in order to inform more tailored interventions, was also noted as a major gap. This study, therefore, uniquely mapped HIV testing uptake at granular sub-location level geographic units, and further mapped clients who were not found at home and those who declined testing.

# CHAPTER THREE
# METHODOLOGY

## 3.1 Introduction to methods

This is an implementation science study. Implementation science is defined as the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine health care to improve the quality and effectiveness of health services (Eccles & Mittman, 2006). There are three overarching aims to the use of implementation science that have been delineated: to describe and/or guide the process of translating research into practice; to understand and/or explain what influences implementation outcomes; and to evaluate implementation (Nilsen, 2020).

This study evaluated strategies that influence an implementation outcome. The outcome was the identification of sub-populations and granular-geographic areas with higher HIV positive yield to inform targeting of HIV testing. Three strategies were evaluated: the use of a HIV predictive algorithm, geospatial analysis of new HIV diagnoses, and mapping of HIV testing uptake.

This study used a retrospective cohort study design. Retrospective cohort studies are a type of observational study, often used in fields related to medicine to study the effect of exposures on health outcomes. Retrospective cohort studies are often quantitative, and use pre-existing secondary research data, such as existing medical records or databases, to identify a group of people with an exposure or risk factor in common. Retrospective cohort studies look backwards in time to examine the relationship between the exposure and the outcome (Johnson, 2018; Talari & Goyal, 2020).

For this study's objective 1, a hospital-based retrospective cohort study design was used. Data collected to primarily document routine provision of HIV testing services, were secondarily used

to develop and validate a HIV predictive risk-score screening algorithm. The data were for patients attending outpatient services, and tested for HIV between September 2017 and May 2018 at six high-volume health facilities in the three counties of Siaya, Kisumu and Homa Bay. For this study's objectives 2 and 3, a community-based retrospective cohort study design was used. Data collected to primarily document routine provision of home-based HIV testing in Siaya County, were secondarily used, in combination with spatial data for administrative unit boundaries and ecological features accessed from DIVA-GIS (https://www.diva-gis.org/gdata), to conduct geospatial analysis of new HIV diagnoses and mapping of HIV testing uptake. The data were for clients accessing community home-based HIV testing in Siaya County between May 2016 to July 2017.

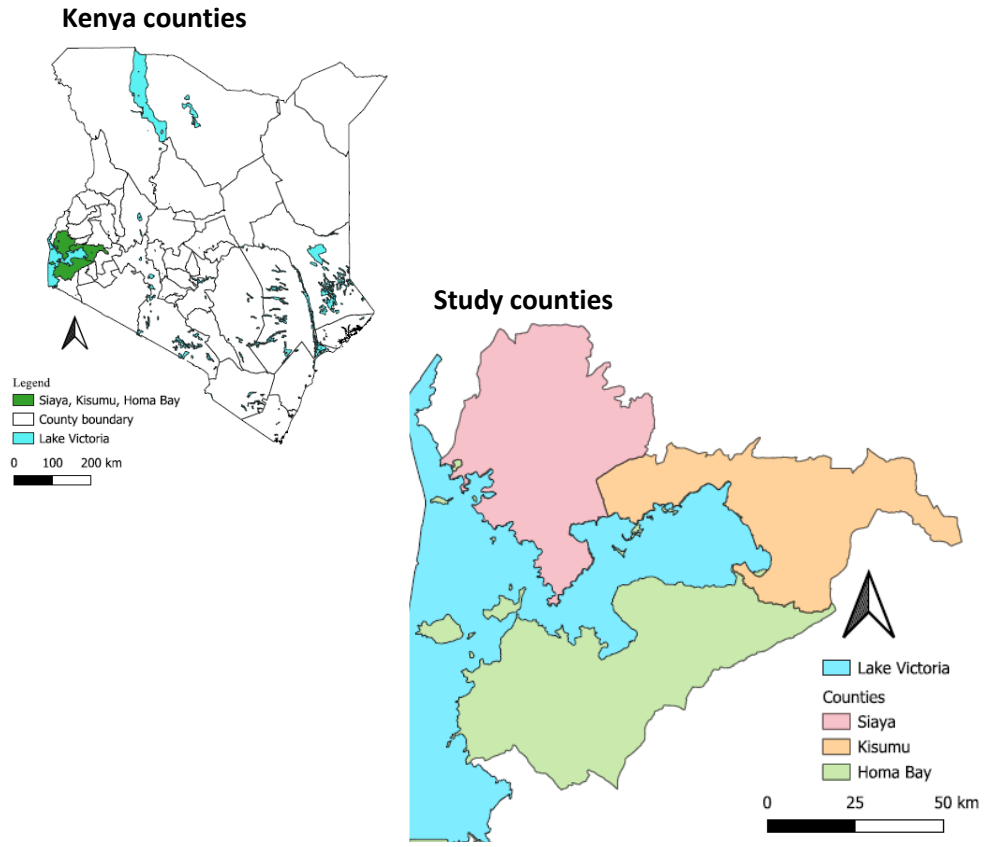
This study was conducted in the three counties of Siaya, Kisumu, and Homa Bay. These three counties have the highest HIV prevalence in Kenya. For study objective 1, the three counties were used; purposively selecting in each county health facilities with the highest outpatient workload and the highest number of patients tested for HIV. For study objectives 2 and 3, Siaya County was used. Although, for objective 2 and 3, it would have been possible to use home-based HIV testing data from the three counties, one county was chosen as the estimated sample of >352,000 clients enumerated for home-based testing was sufficient to demonstrate the mediating effect of geospatial analysis in identifying granular sub-location clusters of new HIV diagnoses. Siaya was chosen, since at the time of this study home-based HIV testing had been completed, and therefore, data was available for secondary analysis in a retrospective cohort design. Although Siaya's home-based HIV testing data was used, geospatial analysis can be done using data from other counties and also, other HIV testing programs.

The counties of Homa Bay, Siaya and Kisumu in the western region of Kenya border Lake Victoria (Figure 3.1), have a total population of 3,280,707, an HIV prevalence ranging between 14% to18%, and an HIV incidence ranging between 5.1 to 6.7 per 1,000 population. Combined, the three counties have approximately 372,000 people living with HIV (Kenya National AIDS and STI Control Programme, 2020a). These counties therefore bear the brunt of HIV in Kenya, having the highest HIV prevalence and incidence. Achieving HIV epidemic control in these three counties is therefore important for Kenya.

Homa Bay County, located in the western region of Kenya (Figure 3.1), has a population of 1,131,950 (Kenya National Bureau of Statistics, 2019) and an area of 3,154 kilometers$^2$ (Kenya State Department for Devolution, n.d). The county is mostly rural, but has several urban centers. The main economic activities are fishing and crop farming, including sugarcane, maize, and sweet potatoes. Homa Bay has a HIV prevalence of 17.9%, the highest among the 47 counties in Kenya (Kenya National AIDS and STI Control Programme, 2020). The county has a HIV incidence of 6.3 per 1,000 population; and estimated 137,482 people living with HIV (Kenya National AIDS and STI Control Programme, 2020).

Kisumu County, located in the western region of Kenya (Figure 3.1), hosts the third largest city in Kenya. It has a population of 1,155,574 (Kenya National Bureau of Statistics, 2019), and a land area of 2,085 kilometers$^2$ (Kenya State Department for Devolution, n.d). The main economic activities are fishing; agriculture including subsistence farming, rice and sugar cane growing; industry and service as the city is a major economic hub for the region; and tourist. Kisumu has a HIV prevalence of 15.8%, the second highest among the 47 counties in Kenya. The county has a HIV incidence of 6.7 per 1,000 population; and an estimated 131,169 people living with HIV (Kenya National AIDS and STI Control Programme, 2020).

Siaya County is also located in the western region of Kenya (Figure 3.1), and has a population of 993,183 (Kenya National Bureau of Statistics, 2019) and a land area of 2,496 kilometers$^2$ (Kenya State Department for Devolution, n.d). The main economic activities are fishing; agriculture including subsistence farming, livestock keeping and rice farming; and small-scale trade. Siaya has a HIV prevalence of 14.4%, the third highest among the 47 counties in Kenya. The county has a HIV incidence of 5.1 per 1,000 population; and an estimated 103,621 people living with HIV (Kenya National AIDS and STI Control Programme, 2020).




Note: the three counties of Siaya, Kisumu and Homa Bay were used for study objective 1; and Siaya County was used for study objectives 2 and 3. Maps generated using a standard Geographical Information System (GIS) program, Quantum GIS version 3.6 (http://qgis.org). The boundary information was obtained as shape files from DIVA-GIS (https://www.diva-gis.org/gdata). Abbreviation: Km- kilometers.

**Figure 3.1: Study area to evaluate a HIV predictive algorithm, geospatial analysis of new HIV diagnoses and mapping of HIV testing uptake in Homa Bay, Kisumu, and Siaya Counties in Kenya**

## 3.2 Methods for study objective 1: The use of a HIV predictive risk-score screening algorithm

This section describes the methods for study objective 1, with the following sub-headings:

- Study design, sites, setting, population, and data management
- Stages in the development and validation of the HIV predictive risk-score algorithm
- Stage one- identifying the need for a HIV predictive algorithm
- Stage two- development of the HIV predictive algorithm according to methodological standards
- Stage three- external validation of the HIV predictive algorithm
- Software used for data analysis.

### 3.2.1 Study design, sites, setting, population, and data management

***Study design***

Using a hospital-based retrospective cohort study design, data collected during the routine provision of HIV testing at five health facilities in the western region of Kenya were used to develop a socio-demographic and behavioral characteristics-based risk-score algorithm to identify sub-populations with higher HIV positive yield to inform targeting of HIV testing. Data from one high-volume facility were used to externally validate the algorithm. Development and validation of the risk-score algorithm followed systematic methodology that has been well described (Laupacis & Sekar, 1997; Moons et al., 2015; Toll, Janssen, Vergouwe, & Moons, 2008; Wasson, Sox, Neff, & Goldman, 1985).

***Study sites***

Seven health facilities that had the highest (1,000–5,000) average monthly outpatient department visits and the highest number of persons aged ≥15 years tested for HIV in the three counties were purposively considered for inclusion in the study (Table 3.1). Although all seven of the selected

health facilities collected HIV behavioral risk data during their routine provision of HIV testing and counseling services, one facility (Bondo County Hospital in Siaya) was found to inconsistently document behavioral risk information and was therefore excluded.

Therefore, data from the following six sites were included in the study: Siaya County Hospital, Homa Bay County Hospital, Mbita County Hospital, Jaramogi Oginga Odinga Referral Hospital, Ahero Sub- County Hospital and Kisumu County Hospital (Table 3.1).

**Table 3.1: Health facilities selected for the evaluation of a sociodemographic and behavioral characteristics-based HIV predictive risk-score screening algorithm in Homa Bay, Kisumu and Siaya Counties**

| County | Facility name[a] | Monthly OPD workload[b] | Number of persons ≥15 years tested for HIV in OPD in 2017[c] | Data used for algorithm development and validation |
|---|---|---|---|---|
| Siaya | Siaya County Hospital | 1,001 | 25,127 | Algorithm development |
| Homa Bay | Homa Bay County Hospital | 895 | 22,771 | |
| Homa Bay | Mbita County Hospital | 2,222 | 12,686 | |
| Kisumu | Jaramogi Oginga Odinga Referral Hospital | 1,552 | 28,580 | |
| Kisumu | Ahero Sub- County Hospital | 2,205 | 20,809 | |
| Kisumu | Kisumu County Hospital | 5,469 | 38,170 | External Validation |
| | **Total** | **13,344** | **148,143** | |

[a]Bondo County Hospital in Siaya, with a monthly OPD workload of 2,906 and 36,738 persons ≥15 years tested for HIV in the OPD in 2017, was found to inconsistently document behavioral risk information and was therefore excluded. [b]Figures reported in Implementing Partner Monthly Acceleration Report, November 2016. [c]PEPFAR Annual Report, 2017. Abbreviations: OPD- outpatient department.

***Study setting and HIV testing procedures***

The six health facilities included in the study offered provider-initiated HIV testing and counseling to out-patients using an opt-out approach. This included screening for HIV-testing eligibility, and provision of pre-test counseling, testing, and post-test counseling to eligible

clients. Eligibility for HIV testing was based on the 2015 Kenya Ministry of Health HIV testing guidelines (Kenya Ministry of Health, 2015), which recommend routine HIV testing to all clients attending health facilities, with annual retesting or more frequently based on HIV-exposure or risk. Based on these guidelines, the following individuals were offered testing: those who have never been tested for HIV; individuals who's last reported negative HIV test result was more than 12 months ago, or who do not know the date of their most recent HIV test; individuals who have signs, symptoms, or a diagnosis of tuberculosis or STI; and those who report recent HIV exposure. In March 2017, eligibility for HIV testing was expanded to increase access to HIV testing services. The expanded eligibility criteria included individuals reporting a negative HIV test result in the past 3 to 12 months, and those reporting a negative HIV test result in the past <3 months, but for whom the test result could not be confirmed in clinic records. Eligible patients were tested for HIV according to the Ministry of Health guidelines using Determine™ and First Response™ rapid point-of-care kits; an individual was considered HIV-negative (uninfected) if the Determine test result was negative, HIV positive (infected) if the Determine and First Response serial test results were positive, and inconclusive if the Determine result was positive and the First Response result was negative. From September 2017, the health facilities in this study used standardized forms to document behavioral risk characteristics routinely assessed by HIV-testing counselors to guide HIV prevention counseling during pre-test counseling sessions.

***Study population***

The study analysis included data from clients aged 15 years and older who were tested for HIV between September 2017 and May 2018 in the outpatient departments of the six study sites, and who had documentation of one or more behavioral risk characteristics. Records for patients with inconclusive HIV test results were excluded. At the six health facilities, data for an entire month

were excluded if ≥50% of patients tested for HIV in that month did not have any documentation of behavioral risk characteristics.

***Study inclusion criteria:***

i. Persons≥15 years of age, attending out-patient services and tested for HIV
ii. Had documentation of one or more behavioral risk characteristics

***Study exclusion criteria:***

i. Pregnant women attending antenatal clinic services
ii. Individuals with no documentation of any behavioral risk characteristics
iii. Patients with inconclusive HIV test results
iv. At the facility level, data for an entire month were excluded if ≥50% of patients tested for HIV in that month did not have any documentation of behavioral risk characteristics

***Data management***

Socio-demographic, HIV screening and testing, and behavioral risk information were collected by lay HIV testing counselors, and recorded manually on standardized forms (Appendix 1) and Ministry of Health registers (Appendix 2). At each health facility, the data were entered into a secure password-protected database. Data meeting the study inclusion criteria were stripped of all identifiers (names and unique patient numbers), assigned new evaluation-specific identification numbers, entered into a study-specific secure password-protected database, and encrypted. Encrypted de-identified data were uploaded from each facility to a central database for study analysis.

### 3.2.2 Stages in the development and validation of a HIV predictive risk-score algorithm

The HIV predictive algorithm development and validation followed standard methodology that has been well documented and described in the literature (Laupacis & Sekar, 1997; Moons et al.,

2015; Toll et al., 2008; Wasson et al., 1985). Three stages of algorithm development and validation were followed: Stage one-identifying the need for a HIV predictive algorithm; Stage two- algorithm development; and Stage three- algorithm external validation.

The next section describes each of these stages in detail.

#### 3.2.2.1 Stage one- identifying the need for a HIV predictive algorithm

A literature and program review were conducted to determine the need and relevance of a HIV predictive algorithm to screen clients attending out-patient services. The program was implementing HIV testing in a universal manner, according to the 2015 Kenya HIV testing, leading to many patients being tested, and few identified HIV positive. There was, therefore, a need to better target HIV testing for efficiency. On the other hand, in the literature review, although many studies had evaluated HIV predictive algorithms among different sub-populations and settings, no algorithms were found, that had been evaluated in the general outpatient setting in sub-Sahara Africa.

#### 3.2.2.2 Stage two- risk-score algorithm development

***Study sites for algorithm development***

Outpatient HIV testing data for persons ≥15 years of age from five of the six health facilities included in this study were used to develop overall and gender-specific risk-score algorithms; two facilities were in Kisumu County (a referral hospital and sub-county hospital), two were in Homa Bay county (a county and sub-county hospital), and one was in Siaya county (a county hospital).

***Primary and predictive outcomes***

The primary outcome in this analysis was an HIV positive result. Socio-demographic and behavioral characteristics were considered for inclusion in the development of the predictive

model if they were among those collected during the routine provision of HIV testing services (collected during the pre-test counseling phase of HIV testing), and have been shown (Dube, Marshall, Ryan, & Omonijo, 2018; Gerbert, Bronstone, McPhee, Pantilat, & Allerton, 1998; Haukoos et al., 2012; Lazzarin, Saracco, Musicco, & Nicolosi, 1991; Oluoch et al., 2011) or hypothesized to be associated with HIV infection. As shown in Table 3.2, these included: socio-demographic characteristics (sex, age, marital status and occupation); behavioral characteristics (change in sexual partners, number of sexual partners, consistent condom use, had sex in exchange for money/favors, engaged in sex work, men who reported having sex with men, female anal sex, injecting drugs for pleasure, had sex under the influence of alcohol or other substance, and coerced to have sex); reported treatment for STI; circumcision status; and specific reasons for HIV testing eligibility (never tested for HIV, interval since last HIV-negative test, having tuberculosis, having an STI, and reporting recent HIV exposure) (Table 3.2). Characteristics such as education level, having an HIV infected sexual partner (Dube et al., 2018) and involvement in fish trade (Kenya National AIDS Control Council, 2009), which have been shown to be associated with HIV infection in other studies, were not collected during the routine provision of HIV testing services.

HIV-risk and behavioral data were collected by trained counselors at a private space, to facilitate patient privacy and reduce social desirability bias.

**Table 3.2: List of socio-demographic and behavioral questions collected during the routine provision of HIV testing services at the study sites in Homa Bay, Kisumu and Siaya Counties**

| |
|---|
| Age and Sex |
| Occupation |
| Marital status |
| Living or staying with spouse |
| Having sex in the last 12 months |
| Number of sexual partners in the last 12 months |
| Any changes in sexual partners in the last 12 months |
| Having been coerced to have sex against ones will in the last 12 months |
| Consistent use of condom in the last 12 months with sexual partner |
| Having sex under the influence of alcohol in the last 12 months |
| Having sex in exchange of money or other favors in the last 12 months |
| Engaging in sex work, men having sex with men or injecting drugs in the last 12 months |
| Been treated for a sexually transmitted infection in the last 12 months |
| Been circumcised (for male) |

Guidelines for sample size estimation for logistic regression based on the concept of event per predictor variable (EPV) have been well published, and were used (Harrell, Frank, Lee, & Mark, 1996; Laupacis & Sekar, 1997; Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). Event refers to the outcome of interest, and predictors refer to the variables whose association with the outcome were assessed.

The validity of the logistic model becomes problematic when the ratio of the numbers of events per variable analyzed is small (Peduzzi et al., 1996). The main concerns have been the accuracy and precision of the regression coefficients, and potentially misleading associations. Three types

of errors have been discussed: over fitting (Type I error) occurs when too many variables in comparison to the number of outcomes, some of which may be "noise," are selected for retention in the final model; under fitting (Type II error) occurs when important variables are omitted from the final model; and paradoxical fitting (Type III error) is produced when a particular factor is given an incorrect direction of association, which is the opposite of the true effect (Peduzzi et al., 1996). It's been noted that using EPV values of less than 10may lead to several problems: the regression coefficients may be biased in both positive and negative directions; the large sample variance estimates from the logistic model may both overestimate and underestimate the sample variance of the regression coefficients; the 90% confidence limits about the estimated values may not have proper coverage; and paradoxical associations (significance in the wrong direction) may be increased (Peduzzi et al., 1996).

Therefore, as a general rule, multivariable logistic regression requires an event-to-predictor ratio of 10:1. In this study, the event was a HIV positive diagnosis; and predictors were the variables whose association with a HIV positive diagnosis were assessed.

The following sixteen predictor variables were considered for inclusion in the multivariable logistic regression model, as predictors of HIV infection: ages 35–39 and 40–44 years; male; manual/domestic or trade/sales/service occupation; married polygamous, widowed or separated/divorced; having >2 sexual partners, a new sexual partner, divorced/separated or widowed in prior 12 months; coerced to have sex or had sex in exchange for money/favors in prior 12 months; reported treatment for STI in prior 12 months; and never been tested for HIV or had a HIV negative result >12 months ago.

Sample size computation:

- The total number of predictor variables in the study, that were included in the multivariable model,were16.
- For the algorithm development sample, to meet the event-to-predictor ratio of 10:1 for the 16 candidate predictor variables, it was estimated that the study would require at least 160 HIV positive patients (derived by multiplying 16 by 10).
- The study used data from five health facilities to develop the algorithm. These were Siaya County Hospital, Homa Bay County Hospital, Mbita County Hospital, Jaramogi Oginga Odinga Referral Hospital, Ahero Sub- County Hospital, and Kisumu County Hospital.
- For the five study sites, 27,692 patients attended out-patient services during the months whose data were included in the study. Of these, 27,685 were screened for HIV testing eligibility, and 24,966 were eligible for testing. A total of 21,764 were tested for HIV, of whom 21,745 had valid HIV test results. A total of 19,458 had behavioral characteristics documented and were included in the study analysis, and 210 were HIV positive. The proportion of patients who were HIV positive among those included in the study and with valid test results was 1.08% (210/19,458).
- The estimated total study sample size required was derived as follows:
    - In order to have 160 HIV positive patients in the sample (to meet the 10:1 event-to-predictor ratio), since the HIV prevalence was 1.08%, this would require a total of 14,815 patients. This is computed as: (160*100%)/1.08% =14,815.
    - Therefore, the estimated sample size for algorithm development was= **14,815.**

***Data analysis for the development and internal validation of the HIV predictive risk-score algorithm***

Patient socio-demographic and behavioral characteristics were summarized using frequencies, percentages, medians, and interquartile ranges. Development of the HIV infection predictive model was conducted in a systematic fashion, using univariable and multivariable analyses.

***Dealing with continuous variables***

It is recommended that in the analysis of predictive algorithms, continuous variables are assessed, and categories developed based on the association of the continuous independent variable with the dependent variable. This is usually done in a generalized additive model.

Based on this, the following was done for this study:

- The association between age and HIV infection was assessed using a generalized additive model, which is the predicted odds of HIV-positivity by age. This informed age categorization into 5-year bands, as shown in Figure 3.2.

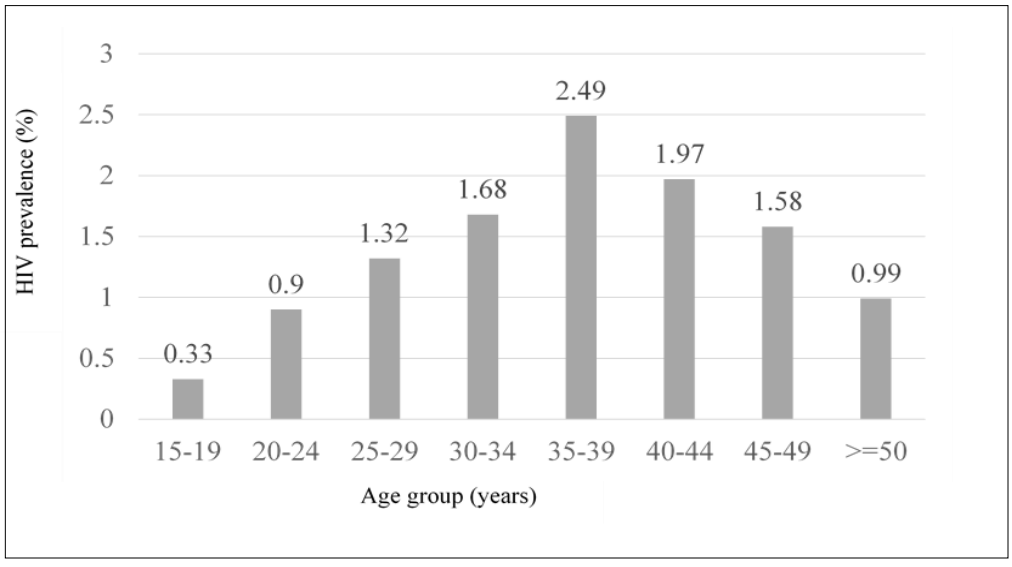



**Figure 3.2: Distribution of HIV prevalence by age in study to evaluate HIV predictive algorithm**

- The age-bands were further categorized into groups according to their HIV prevalence (the proportion of HIV infected individuals), as shown in Table 3.3. In doing this, 5-year age bands that had HIV prevalence that were close to each other were grouped together.

**Table 3.3: Age categorization in study to evaluate HIV predictive algorithm**

| Age (years) | HIV prevalence | Age category |
|---|---|---|
| 15-19 | 0.33% | Age category 1 |
| 20-24 | 0.90% | Age category 1 |
| 25-29 | 1.32% | Age category 2 |
| 30-34 | 1.68% | Age category 2 |
| 35-39 | 2.49% | Age category 3 |
| 40-44 | 1.97% | Age category 3 |
| 45-49 | 1.58% | Age category 2 |
| ≥50 | 0.99% | Age category 1 |

- The following age-groups were therefore created:
  - Ages 15–19, 20–24 and ≥50 years (HIV prevalence range of 0.33%–0.99%).
  - Ages 25–29, 30–34 and 45–49 years (HIV prevalence range of 1.32%–1.68%).
  - Ages 35–39 and 40–44 years (HIV prevalence range of 1.97%–2.49%).

***Dealing with missing data***

As missing data may cause problems in regression analysis, for this study, missing data were excluded from univariable and multivariable analyses.

***Univariable analysis***

Univariable analysis was conducted to assess the independent association between the socio-demographic and behavioral characteristics and HIV infection, by computing odds ratios (ORs) and their corresponding 95% confidence intervals (CIs) and p values (significant at $p \leq 0.05$). Two variables were not included in the univariable analysis: having sex in the prior 12 months, as multiple characteristics were assessed only for those who had sex in the prior 12 months, and

consistent condom use with a sexual partner, as the documentation format made this variable difficult to interpret.

***Multivariable analysis and best HIV predictive model selection***

The initial full multivariable analysis included all variables with a significantly higher odds (OR >1.0) of HIV infection in univariable analysis, and those selected based on prior knowledge of an association with HIV infection. The variables in the full multivariable analysis were evaluated in a stepwise multivariable logistic regression that incorporated Akaike information criterion for model selection, to identify the model/algorithm that best predicted HIV infection. Corresponding ORs, beta regression coefficients, and 95% CIs were computed.

***HIV predictive algorithm internal validation***

The final model was internally validated using 10-fold cross-validation. In the 10-fold cross-validation, the original sample was randomly partitioned into 10 equal size sub-samples. The cross-validation process was then repeated 10 times, with each of the 10 sub-samples used once as the validation dataset. The results from the 10 repeated sub-samples were then combined to produce a single estimation.

***Assessment of the performance of the algorithm***

The ability of the final risk-score algorithm to discriminate between individuals with, and without, HIV infection was evaluated by computing the average area under the receiver operating curve (AUC, the area under a plot of sensitivity and the inverse of specificity) from the ten different cross-validation models. R-squared ($R^2$) was computed to assess the extent to which the HIV prevalence variability can be explained by the model.

***Creation of HIV risk-scores and risk-score categories***

Risk-scores for each variable in the final model were created by multiplying the corresponding beta regression coefficient by 10 and rounding to the nearest integer for ease of calculation. Each patient's total risk-score was generated by summing the scores for all variables met.

To create risk-score categories, patient risk-scores were arranged in ascending order. The corresponding HIV prevalence for patients meeting each score was computed and used to identify mutually exclusive cut-points for unique risk-score groupings. The aggregate HIV prevalence and corresponding CIs were then calculated for each defined risk-score grouping.

**3.2.2.3 Stage three- HIV predictive risk-score algorithm external validation**

External validation is the action of testing the original prediction algorithm in a set of new patients to determine whether the algorithm works to a satisfactory degree. External validation is necessary to assess an algorithm's reproducibility and generalizability (Debray, Vergouwe, Koffijberg, Nieboer, Steyerberg, & Moons, 2015; Ramspek, Jager, Dekker, Zoccali, & van Diepen, 2021).

***Study site for risk-score algorithm external validation***

Data from Kisumu County Hospital, a health facility among the six high-volume sites selected for inclusion in this study, were used to externally validate the overall and the gender-specific risk-score algorithms developed. This hospital had the highest outpatient workload and number (~38,000) of persons ≥15 years tested for HIV in the outpatient department in 2017, in the three counties of Siaya, Kisumu and Homa Bay. Although the hospital was located in the same region where the algorithm was developed, the hospital's population of outpatient attendees had major

differences with that of the five facilities used for algorithm development, as highlighted in Section 4.2.3, and Table 4.5.

These population differences gave strength to the algorithm's validation process and performance assessment for reproducibility. Additionally, it gave strength for generalizability, that would apply to several counties in the western region of Kenya, that have fairly similar population characteristics. It is however recommended that, before use in an entirely different region, the algorithm should first be validated in that region. Procedures for HIV testing, documentation of socio-demographic and behavioral characteristics, and management of HIV testing data were similar to those earlier described for the other facilities included in the study.

***Sample size calculation for external validation of HIV predictive risk-score algorithm***

Similar to the algorithm development sample, the guidelines of sample size estimation for logistic regression based on the concept of event per predictor variable (EPV) were used to compute the sample size required for external validation (Harrell et al., 1996; Laupacis & Sekar, 1997; Peduzzi et al., 1996). The general rule of the event-to-predictor ratio of 10:1 was used.

Sample size computation:

- The total number of predictor variables in the study, that were included in the multivariable model were 16.
- For the algorithm external validation sample, to meet the event-to-predictor ratio of 10:1 for the 16 candidate predictor variables, it was estimated that the study would require at least 160 HIV positive patients (derived by multiplying 16 by 10).
- The study used data from one health facility to externally validate the algorithm. This was Kisumu County Hospital.

- For the one study site used for external validation, 20,055 patients attended out-patient services during the months included in the study. Of these, 19,415 were screened for HIV testing eligibility, and 13,213 were eligible for testing. A total of 13,012 were tested for HIV, of whom 13,010 had valid HIV test results. A total of 11,330 had behavioral characteristics documented and were included in the study analysis, and 174 were HIV positive. The proportion of HIV positive patients among those included in the study with valid test results was 1.54% (174/11,330).
- The estimated total study sample size required was derived as follows:
    - In order to have 160 HIV positive patients in the sample (to meet the 10:1 event-to-predictor ratio), since the HIV prevalence was 1.54%, this would require 10,390 patients. This was computed as: (160*100%)/1.54% =10,390.
    - Therefore, the estimated sample size for the algorithm external validation was = **10,390.**

***Data analysis for the external validation of the HIV predictive risk-score algorithm***

For external validation, each patient's risk-score was generated using the risk-score algorithm developed, and patients were grouped into respective risk-score categories. HIV prevalence and corresponding CIs for each risk-score category were then calculated. The AUC and $R^2$ were computed in order to assess the algorithm's discrimination performance, and the extent to which variability in HIV prevalence is explained by the model, respectively.

### 3.2.3 Software for data analysis

Data were managed using Stata Statistical Software version 14 (StataCorp, College Station, TX). The Classification And REgression Training (CARET) package in R version 3.6.2 (R Core Team, 2013) was used for predictive modeling to perform 10-fold cross-validation, and to assess

the performance of the predictive algorithm by computing the AUC and the $R^2$. The CARET package found in R is a front-end package that wraps around a lot of the prediction algorithms and tools in the R programming language, and is used for complex regression functions.

## 3.3 Methods for study objective 2: The use of geospatial analysis of new HIV diagnoses in identifying areas with higher HIV positive yield

This section describes the methods for study objective 2, under the following sub-headings:

- Study design, sites, setting, population, and data management
- Data analysis

### 3.3.1 Study design, area, procedures for home-based HIV testing, study population, sample size, data management

***Study design***

Using a community-based retrospective cohort study design, this study used data from community home-based HIV testing conducted as part of a routine public health program in Siaya County, western Kenya. Geospatial analysis was used to assess geographic clusters of new HIV diagnoses, and a spatially integrated Bayesian model to describe factors associated with new HIV diagnoses in order to inform the targeting of HIV interventions to finer geographic areas and sub-populations. Home-based testing was supported by the United States President's Emergency Plan for AIDS Relief (PEPFAR) through the United States Centers for Disease Control and Prevention (CDC), under the Impact Research and Development Organization cooperative agreement.

***Study area***

Siaya County borders Lake Victoria in western Kenya. The population is predominantly rural and includes fishing communities living along the lake's beaches. Administratively, the county

consists of six sub-counties, which are subdivided into 30 wards, and further into 179 sub-locations, and 2,285 villages. In 2016 and 2017 intensified routine HIV testing was implemented in Siaya, and included biannual testing offered to fishing communities living along the beaches, and home-based testing offered to inland residents of the county.

***Procedures for home-based HIV testing***

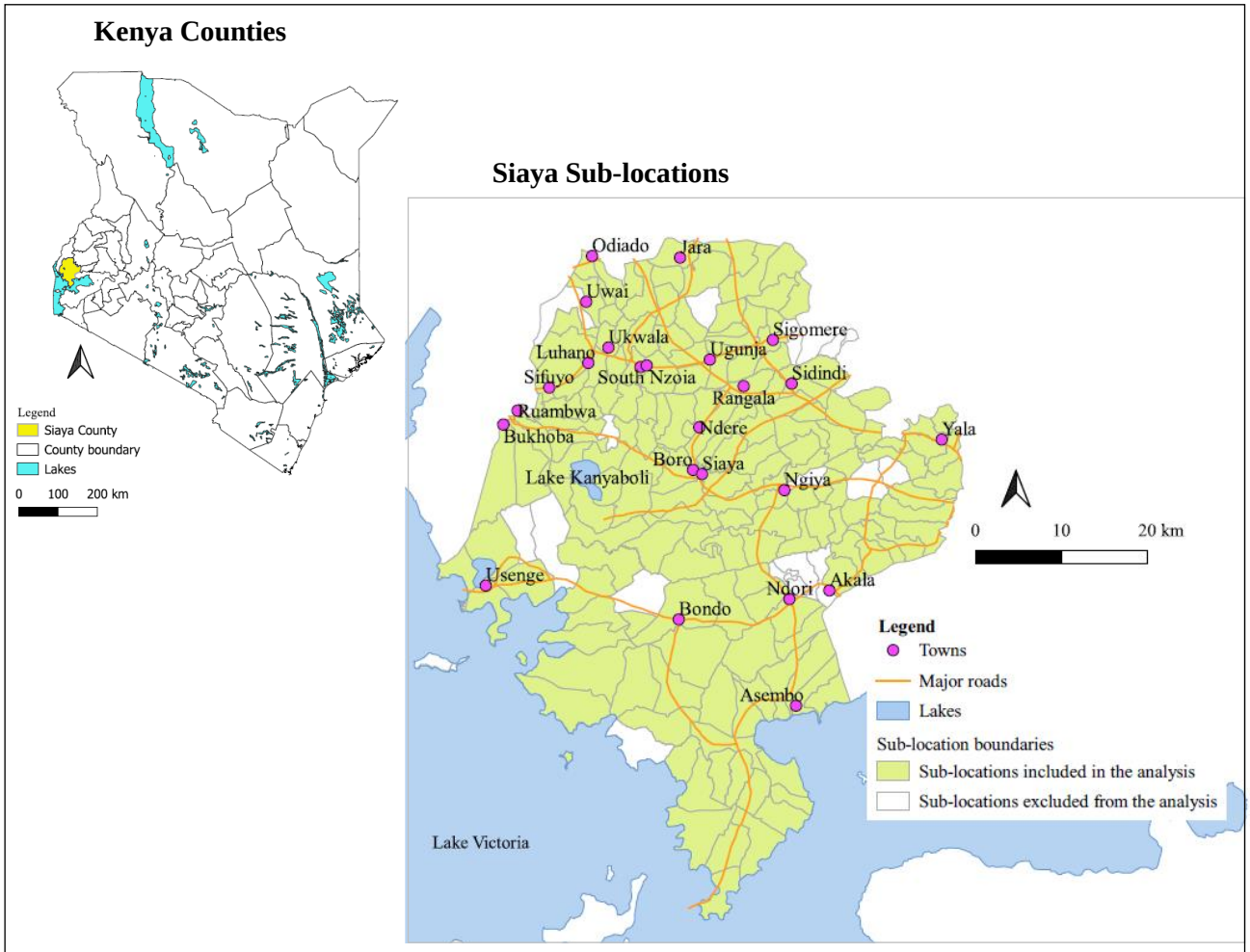
For home-based HIV testing, all households in the inland geographic areas were visited to enumerate occupants and assess their eligibility for HIV testing. Household occupants were enumerated if they would be resident in the household for one or more months following enumeration. Clients aged ≥15 years were eligible for HIV testing if they reported having never been tested for HIV; reported a negative HIV test done more than three months ago; had signs, symptoms or a diagnosis of tuberculosis, or a sexually transmitted infection; or reported a recent (within 3 months) HIV exposure such as unprotected sex with a partner of unknown or positive HIV status. Children aged 14 years and below were eligible for testing if their biological mother was known to be HIV infected or deceased. Within one month of enumeration, trained lay counselors offered pre-test counseling, HIV testing and post-test counseling to those eligible. Counselors made up to three follow-up visits to offer testing to those not found at home. HIV testing was offered according to the 2015 Kenya HIV testing guidelines (Kenya Ministry of Health, 2015) using Determine™ (Alere Medical Co. Ltd, 2015) and First Response® (Premier Medical Corporation Limited) rapid point of care kits[1]. An individual was considered HIV-negative (uninfected) if the Determine test result was negative (considered a conclusive negative result), HIV positive (infected) if both the Determine and First Response serial tests results were positive (considered a conclusive positive result), and inconclusive if the Determine test was

[1]From the manufacturer's package insert, Determine™ has a sensitivity of 100% and specificity of 99.8%; while First Response® has a sensitivity of 100% and specificity of 99.5%.

positive and First Response test was negative. Clients with inconclusive HIV test results were referred to a health facility for follow-up testing according to Kenya Ministry of Health guidelines.

***Study population and inclusion criteria for home-based HIV testing data***

Data for clients aged ≥15 years who received routine home-based HIV testing in Siaya County from May 2016 to July 2017 were retrospectively analyzed. Home-based testing data for children aged <15 years, and data collected as part of biannual HIV testing of fishing communities, were excluded from the analysis. Data were spatially analyzed at the sub-location level; sub-locations in which all, or more than half of households, were enumerated, were included in the analysis. Out of the 179 sub-locations in the county (Appendix 3), data from 161 sub-locations met the criteria for inclusion (156 sub-locations in which all households were enumerated, and 5 in which ≥50% of households were enumerated). Figure 3.3 shows the sub-locations included in the study analysis.




Note: sub-locations included in the analysis were 161 that met the inclusion criteria (156 sub-locations in which all households were enumerated, and 5 in which ≥50% of households were enumerated). A total of 18 sub-locations not meeting inclusion criteria (<50% of households were enumerated) were excluded. Sub-locations were mapped using a standard Geographical Information System (GIS) program, Quantum GIS version 3.6 (http://qgis.org). The boundary information for sub-locations in Siaya was obtained as shapefiles from DIVA-GIS (https://www.diva-gis.org/gdata). Abbreviation: Km- kilometers.

**Figure 3.3: Siaya County sub-locations included in the analysis of study to evaluate HIV predictive algorithm**

### *Sample size*

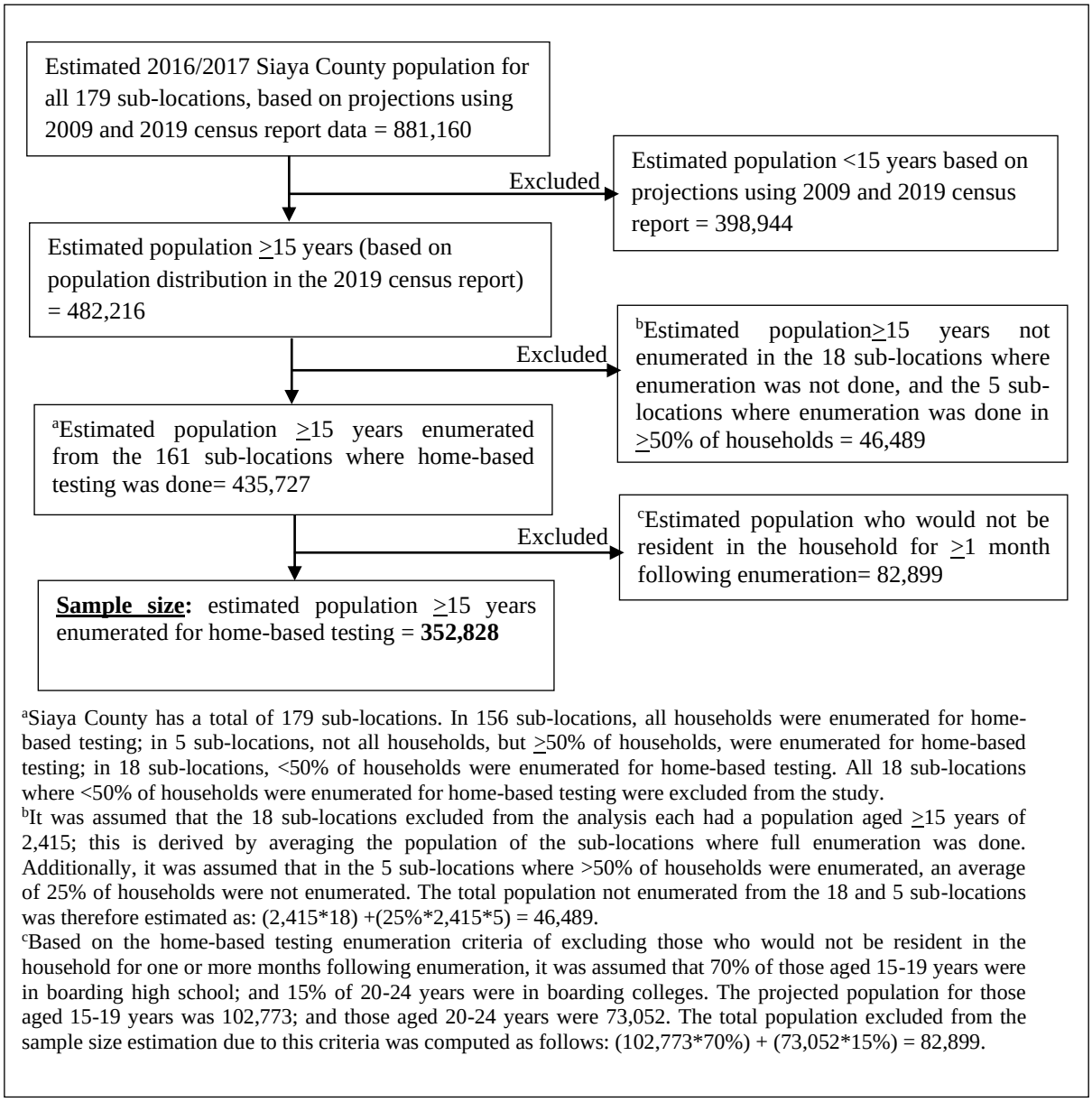
Home-based HIV testing that was done in Siaya County was offered through a house-to-house approach, ensuring all households in an area were enumerated and testing offered to household members. This study used data for clients offered home-based HIV testing, in order to enable

population-level geospatial analysis and mapping of clusters of new HIV diagnoses in Siaya County. The study sample therefore followed the Census Method, which is also known as the Complete Enumeration Method (Chan, McGarey, & Sclafani, 2018; Kish, 1979). All the clients who were enumerated for routine home-based HIV testing were considered for inclusion in the study. As the program aimed to offer home-based HIV testing to all the sub-locations in Siaya County, the study sample size was therefore computed as the projected estimated population for Siaya County, based on the most recent census data, and excluded clients who did not meet the study or home-based testing inclusion criteria.

Clients enumerated for home-based HIV testing were persons 15 years or more of age, who would be resident in Siaya for more than one month following enumeration. The following were excluded from the study: sub-locations where ≥50% of households were not enumerated for home-based testing; and clients who would not be resident in Siaya for one month or more following enumeration.

Figure 3.4 describes how the population sample enumerated for home-based testing was estimated. In 2016/2017, Siaya County had an estimated total population of ~881,144, projected using 2009 (Kenya National Bureau of Statistics, 2009) and 2019 (Kenya National Bureau of Statistics, 2019) Kenya population census reports. Based on the estimated population of individuals aged ≥15 years (482,216), and excluding those in sub-locations where home-based testing enumeration was not done (46,489) and those who would not be resident in the household for one or more months following enumeration (82,899), the study sample size was therefore estimated as 352,000 individuals aged ≥15 years who would be enumerated for home-based testing.

Table 3.4 shows the population enumerated and offered testing by sub-location, for the 161 sub-locations included in the study. Overall, 365,798 clients were enumerated for home-based testing.




[a]Siaya County has a total of 179 sub-locations. In 156 sub-locations, all households were enumerated for home-based testing; in 5 sub-locations, not all households, but ≥50% of households, were enumerated for home-based testing; in 18 sub-locations, <50% of households were enumerated for home-based testing. All 18 sub-locations where <50% of households were enumerated for home-based testing were excluded from the study.

[b]It was assumed that the 18 sub-locations excluded from the analysis each had a population aged ≥15 years of 2,415; this is derived by averaging the population of the sub-locations where full enumeration was done. Additionally, it was assumed that in the 5 sub-locations where >50% of households were enumerated, an average of 25% of households were not enumerated. The total population not enumerated from the 18 and 5 sub-locations was therefore estimated as: (2,415*18) +(25%*2,415*5) = 46,489.

[c]Based on the home-based testing enumeration criteria of excluding those who would not be resident in the household for one or more months following enumeration, it was assumed that 70% of those aged 15-19 years were in boarding high school; and 15% of 20-24 years were in boarding colleges. The projected population for those aged 15-19 years was 102,773; and those aged 20-24 years were 73,052. The total population excluded from the sample size estimation due to this criteria was computed as follows: (102,773*70%) + (73,052*15%) = 82,899.

**Figure 3.4: Sample size calculated for home-based testing enumeration in Siaya County, May 2016 to July 2017**

**Table 3.4: Population enumerated and offered home-based testing in the 161 sub-locations in Siaya, May 2016 to July 2017**

| Sub-location | Number enumerated | Eligibility assessed | Eligible | % Eligible | Tested | % Tested |
|---|---|---|---|---|---|---|
| East Asango | 355 | 333 | 302 | 91% | 246 | 81% |
| Anyiko | 614 | 611 | 484 | 79% | 423 | 87% |
| Simur East | 614 | 613 | 504 | 82% | 434 | 86% |
| Doho West | 654 | 652 | 559 | 86% | 502 | 90% |
| Sifuyo West | 669 | 667 | 568 | 85% | 468 | 82% |
| Doho East | 671 | 667 | 571 | 86% | 505 | 88% |
| Kodiere | 719 | 701 | 640 | 91% | 567 | 89% |
| Sirembe | 750 | 722 | 620 | 86% | 392 | 63% |
| Ojwado 'B' | 763 | 730 | 524 | 72% | 484 | 92% |
| Nguge | 808 | 723 | 610 | 84% | 519 | 85% |
| Sifuyo East | 856 | 847 | 723 | 85% | 652 | 90% |
| Malunga East | 866 | 847 | 721 | 85% | 469 | 65% |
| Ulafu | 892 | 874 | 769 | 88% | 717 | 93% |
| Kalkada Uradi | 898 | 892 | 781 | 88% | 729 | 93% |
| Maranda | 914 | 862 | 728 | 84% | 517 | 71% |
| Ndori | 933 | 903 | 820 | 91% | 495 | 60% |
| Kabura Uhuyi | 970 | 961 | 843 | 88% | 758 | 90% |
| Simur Kondiek | 1,002 | 997 | 790 | 79% | 636 | 81% |
| Kochieng 'B' | 1,020 | 1,001 | 745 | 74% | 665 | 89% |
| Asayi | 1,032 | 1,015 | 907 | 89% | 531 | 59% |
| Kaugagi Hawinga | 1,084 | 1,077 | 903 | 84% | 796 | 88% |
| Wagai East | 1,113 | 1,067 | 985 | 92% | 834 | 85% |
| Bar Olengo | 1,122 | 1,110 | 841 | 76% | 771 | 92% |
| Sigoma Uranga | 1,158 | 1,153 | 1,014 | 88% | 944 | 93% |
| Karadolo East | 1,166 | 1,138 | 909 | 80% | 788 | 87% |
| Wagai West | 1,179 | 1,115 | 979 | 88% | 833 | 85% |
| North Rambula | 1,183 | 1,140 | 960 | 84% | 882 | 92% |
| Kapiyo | 1,189 | 1,174 | 1,013 | 86% | 728 | 72% |
| Rera | 1,199 | 1,188 | 1,045 | 88% | 599 | 57% |
| Utonga | 1,208 | 1,189 | 1,053 | 89% | 807 | 77% |
| Umala | 1,209 | 1,181 | 1,006 | 85% | 889 | 88% |
| Obambo | 1,228 | 1,209 | 982 | 81% | 941 | 96% |
| Olwa | 1,229 | 1,176 | 954 | 81% | 807 | 85% |
| Randago | 1,255 | 1,228 | 1,042 | 85% | 888 | 85% |
| Siriwo | 1,271 | 1,261 | 1,147 | 91% | 769 | 67% |
| Bar Osimbo | 1,298 | 1,256 | 1,067 | 85% | 929 | 87% |
| Sigomre | 1,300 | 1,252 | 1,074 | 86% | 871 | 81% |
| Kochieng 'A' | 1,303 | 1,235 | 1,002 | 81% | 966 | 96% |
| Simenya | 1,323 | 1,300 | 1,056 | 81% | 887 | 84% |
| Simur | 1,324 | 1,316 | 1,098 | 83% | 1,060 | 97% |

| Sub-location | Number enumerated | Eligibility Assessed | Eligible | % Eligible | Tested | % Tested |
|---|---|---|---|---|---|---|
| Komenya Kalaka | 1,331 | 1,325 | 1,208 | 91% | 1,076 | 89% |
| Nyabeda | 1,353 | 1,321 | 1,184 | 90% | 991 | 84% |
| South Rambula | 1,358 | 1,321 | 1,086 | 82% | 873 | 80% |
| Komenya Kowala | 1,376 | 1,352 | 1,246 | 92% | 1,155 | 93% |
| Nyamila | 1,376 | 1,343 | 1,105 | 82% | 984 | 89% |
| Gangu | 1,395 | 1,380 | 1,191 | 86% | 1,117 | 94% |
| Kukumu_Kombewa | 1,446 | 1,390 | 1,115 | 80% | 989 | 89% |
| Masat West | 1,453 | 1,445 | 1,315 | 91% | 1,211 | 92% |
| Dienya West | 1,474 | 1,387 | 1,269 | 91% | 1,106 | 87% |
| Mahaya | 1,499 | 1,474 | 1,262 | 86% | 993 | 79% |
| Masat East | 1,525 | 1,516 | 1,265 | 83% | 1,172 | 93% |
| Onyinyore | 1,571 | 1,418 | 1,282 | 90% | 1,001 | 78% |
| Rageng'ni | 1,572 | 1,464 | 1,289 | 88% | 1,151 | 89% |
| Nyalgunga | 1,583 | 1,549 | 1,322 | 85% | 1,165 | 88% |
| Kaugagi Udenda | 1,613 | 1,602 | 1,404 | 88% | 1,345 | 96% |
| Othach | 1,613 | 1,528 | 1,298 | 85% | 923 | 71% |
| Uriri | 1,616 | 1,559 | 1,393 | 89% | 819 | 59% |
| Rachar | 1,662 | 1,612 | 1,402 | 87% | 1,278 | 91% |
| Pala | 1,666 | 1,455 | 1,202 | 83% | 1,026 | 85% |
| Kandenge | 1,675 | 1,615 | 1,174 | 73% | 1,102 | 94% |
| Malunga West | 1,677 | 1,627 | 1,446 | 89% | 876 | 61% |
| Kathieno 'B' | 1,695 | 1,668 | 1,504 | 90% | 1,016 | 68% |
| Sumba | 1,708 | 1,698 | 1,553 | 91% | 1,413 | 91% |
| Mur_Malanga | 1,726 | 1,688 | 1,466 | 87% | 1,166 | 80% |
| Koyeyo | 1,747 | 1,680 | 1,296 | 77% | 1,213 | 94% |
| South Ramba | 1,770 | 1,745 | 1,518 | 87% | 1,380 | 91% |
| West Asango | 1,809 | 1,793 | 1,570 | 88% | 1,414 | 90% |
| West Migwena | 1,829 | 1,787 | 1,570 | 88% | 1,223 | 78% |
| Mungao | 1,855 | 1,790 | 1,521 | 85% | 1,282 | 84% |
| Yiro East | 1,868 | 1,802 | 1,405 | 78% | 1,186 | 84% |
| Nyadorera 'B' | 1,903 | 1,873 | 1,654 | 88% | 1,451 | 88% |
| Nyamsenda | 1,910 | 1,892 | 1,606 | 85% | 1,447 | 90% |
| Ndere | 1,981 | 1,934 | 1,700 | 88% | 1,363 | 80% |
| Got Ramogi | 2,006 | 1,861 | 1,591 | 85% | 1,355 | 85% |
| Magoya | 2,014 | 1,954 | 1,619 | 83% | 1,441 | 89% |
| Gombe | 2,019 | 1,973 | 1,728 | 88% | 999 | 58% |
| Ligala | 2,029 | 2,017 | 1,699 | 84% | 1,532 | 90% |
| Hono | 2,032 | 2,004 | 1,786 | 89% | 1,676 | 94% |
| East Katwenga | 2,045 | 1,982 | 1,813 | 91% | 1,628 | 90% |

| Sub-location | Number enumerated | Eligibility Assessed | Eligible | % Eligible | Tested | % Tested |
|---|---|---|---|---|---|---|
| Abom | 2,065 | 1,934 | 1,630 | 84% | 1,315 | 81% |
| Pap Oriang | 2,083 | 2,020 | 1,729 | 86% | 1,502 | 87% |
| Kathieno 'C' | 2,089 | 1,998 | 1,841 | 92% | 1,237 | 67% |
| Ngunya | 2,116 | 2,036 | 1,787 | 88% | 1,646 | 92% |
| Jina | 2,117 | 2,058 | 1,865 | 91% | 1,772 | 95% |
| Siranga | 2,143 | 2,125 | 1,856 | 87% | 1,650 | 89% |
| Barding | 2,146 | 1,940 | 1,688 | 87% | 1,465 | 87% |
| Got Osimbo | 2,149 | 2,095 | 1,870 | 89% | 1,486 | 79% |
| Bar-Agulu | 2,193 | 2,159 | 1,844 | 85% | 1,617 | 88% |
| Uyundo | 2,203 | 2,187 | 2,039 | 93% | 1,826 | 90% |
| Nyangoma_Alego | 2,234 | 2,106 | 1,860 | 88% | 1,678 | 90% |
| West Karadolo | 2,239 | 2,220 | 1,897 | 85% | 1,733 | 91% |
| Masumbi | 2,250 | 2,227 | 1,952 | 88% | 1,771 | 91% |
| Usigu | 2,267 | 2,151 | 1,875 | 87% | 1,575 | 84% |
| Tingare West | 2,273 | 2,174 | 1,981 | 91% | 1,691 | 85% |
| Nyalenya | 2,275 | 2,243 | 1,991 | 89% | 1,784 | 90% |
| Komolo | 2,278 | 2,240 | 1,799 | 80% | 1,697 | 94% |
| Lundha | 2,354 | 2,281 | 2,003 | 88% | 1,559 | 78% |
| Ramunde | 2,375 | 2,354 | 2,088 | 89% | 1,323 | 63% |
| Ulamba | 2,383 | 2,272 | 1,968 | 87% | 1,227 | 62% |
| Nyadorera 'A' | 2,411 | 2,391 | 2,103 | 88% | 1,918 | 91% |
| Dienya East | 2,421 | 2,366 | 2,212 | 93% | 2,058 | 93% |
| Nyawara | 2,438 | 2,409 | 2,155 | 89% | 1,370 | 64% |
| Bar Chando | 2,464 | 2,285 | 1,979 | 87% | 1,630 | 82% |
| Anyiko_Yala | 2,468 | 2,417 | 2,286 | 95% | 2,075 | 91% |
| Ojwando 'A' | 2,480 | 2,353 | 1,720 | 73% | 1,529 | 89% |
| Mur_Ngiya | 2,482 | 2,428 | 2,123 | 87% | 1,877 | 88% |
| Ambira | 2,527 | 2,468 | 2,191 | 89% | 1,964 | 90% |
| Kambare | 2,529 | 2,402 | 2,141 | 89% | 1,498 | 70% |
| Yiro West | 2,559 | 2,516 | 1,977 | 79% | 1,694 | 86% |
| Ndenga | 2,571 | 2,537 | 2,038 | 80% | 1,743 | 86% |
| Jera | 2,595 | 2,572 | 2,039 | 79% | 1,772 | 87% |
| Rangala | 2,646 | 2,505 | 2,059 | 82% | 1,776 | 86% |
| Nyajuok | 2,655 | 2,620 | 2,151 | 82% | 1,888 | 88% |
| Kathieno 'A' | 2,663 | 2,604 | 2,378 | 91% | 1,801 | 76% |
| Got Agulu | 2,671 | 2,573 | 2,296 | 89% | 1,701 | 74% |
| Maliera | 2,689 | 2,581 | 2,228 | 86% | 1,740 | 78% |
| Akom | 2,715 | 2,604 | 2,430 | 93% | 2,285 | 94% |
| Tingare East | 2,725 | 2,642 | 2,416 | 91% | 1,913 | 79% |
| Yenga | 2,726 | 2,711 | 2,282 | 84% | 2,221 | 97% |

| Sub-location | Number enumerated | Eligibility Assessed | Eligible | % Eligible | Tested | % Tested |
|---|---|---|---|---|---|---|
| Ligega | 2,738 | 2,665 | 2,238 | 84% | 1,828 | 82% |
| East Migwena | 2,744 | 2,557 | 2,255 | 88% | 2,150 | 95% |
| Got Abiero | 2,750 | 2,642 | 2,283 | 86% | 1,945 | 85% |
| Got Regea | 2,769 | 2,714 | 2,432 | 90% | 2,094 | 86% |
| Nyabera | 2,804 | 2,724 | 2,257 | 83% | 2,079 | 92% |
| Kagwa | 2,839 | 2,787 | 2,401 | 86% | 2,034 | 85% |
| Usenge | 2,853 | 2,729 | 2,398 | 88% | 1,852 | 77% |
| Kagonya | 2,855 | 2,807 | 2,411 | 86% | 2,070 | 86% |
| Omia Diere | 2,908 | 2,894 | 2,647 | 91% | 2,429 | 92% |
| Naya | 2,974 | 2,901 | 2,526 | 87% | 1,951 | 77% |
| Sega | 3,054 | 2,992 | 2,628 | 88% | 2,293 | 87% |
| Kokwiri | 3,156 | 3,113 | 2,780 | 89% | 2,403 | 86% |
| Kobong' | 3,159 | 3,107 | 2,799 | 90% | 2,613 | 93% |
| Omia Mwalo | 3,191 | 3,164 | 2,960 | 94% | 2,773 | 94% |
| Umala_Ugunja | 3,209 | 3,155 | 2,763 | 88% | 2,513 | 91% |
| Lieta | 3,292 | 3,216 | 2,874 | 89% | 2,455 | 85% |
| Lihanda | 3,343 | 3,234 | 2,940 | 91% | 2,778 | 94% |
| Ochieng'a | 3,372 | 3,344 | 3,159 | 94% | 2,926 | 93% |
| Bar Sauri | 3,396 | 3,365 | 3,208 | 95% | 3,020 | 94% |
| Nyamninia | 3,408 | 3,337 | 3,079 | 92% | 2,877 | 93% |
| West Katwenga | 3,413 | 3,318 | 3,054 | 92% | 2,835 | 93% |
| Nyandiwa | 3,533 | 3,456 | 3,161 | 91% | 2,716 | 86% |
| Mulaha | 3,548 | 3,503 | 3,149 | 90% | 2,625 | 83% |
| Uyawi | 3,562 | 3,348 | 2,735 | 82% | 2,304 | 84% |
| Malanga | 3,607 | 3,424 | 2,949 | 86% | 2,685 | 91% |
| Uranga | 3,651 | 3,557 | 3,264 | 92% | 3,022 | 93% |
| Ndigwa | 3,736 | 3,713 | 3,315 | 89% | 3,176 | 96% |
| North Ramba | 3,871 | 3,760 | 3,319 | 88% | 2,877 | 87% |
| Ajigo | 4,109 | 3,912 | 3,531 | 90% | 2,972 | 84% |
| Nyandiwa_Yala | 4,140 | 3,877 | 3,611 | 93% | 3,133 | 87% |
| Omia Malo | 4,266 | 4,227 | 3,950 | 93% | 3,745 | 95% |
| Siger | 4,382 | 4,297 | 3,811 | 89% | 3,107 | 82% |
| Ugunja | 4,428 | 4,315 | 3,753 | 87% | 3,246 | 86% |
| Memba | 4,481 | 4,371 | 4,016 | 92% | 3,561 | 89% |
| Marenyo | 4,534 | 4,374 | 3,940 | 90% | 3,557 | 90% |
| Nyangoma | 4,744 | 4,362 | 3,663 | 84% | 3,009 | 82% |
| Nyagoko | 5,032 | 4,967 | 4,513 | 91% | 4,224 | 94% |
| Ramula | 5,148 | 4,963 | 4,547 | 92% | 4,164 | 92% |
| Karapul | 5,444 | 5,289 | 4,883 | 92% | 4,079 | 84% |
| Masala | 5,685 | 5,613 | 5,263 | 94% | 4,800 | 91% |
| Nyawita | 7,777 | 7,541 | 6,887 | 91% | 5,564 | 81% |
| Bar-Kowino | 10,105 | 9,632 | 8,708 | 90% | 7,816 | 90% |
| **Total** | **365,798** | **355,277** | **312,223** | **88%** | **268,543** | **86%** |

***Data management***

Routine home-based HIV testing data collected included socio-demographic characteristics: age, sex, marital status, and relationship to household head; sub-county, ward, sub-location, and village of residence; and HIV test eligibility criteria and test results. Data collected were manually recorded on standardized enumeration forms (Appendix 4) and the Ministry of Health HIV testing registers (Appendix 5) by lay counselors. At a central (office) location, data clerks reviewed the data for completeness and accuracy, and entered it into a secure password-protected Microsoft Access database.

For this study, data collected during the routine provision of HIV testing services were stripped of all identifiers (names and unique patient numbers) and each record assigned a new study-specific identification number. The analytic dataset was saved in a secure password-protected database.

### 3.3.2 Data analysis

***Descriptive analysis***

Frequencies, proportions, medians, and interquartile ranges were calculated to summarize the data. The proportion of new HIV positive clients (new HIV positive yield) was defined as the total number of clients newly identified HIV positive among those with a conclusive test result. The proportion of total HIV positive clients was calculated as the sum of new HIV positive and previously identified HIV-infected clients among those assessed for HIV test eligibility.

***Spatial data analysis***

Spatial data for administrative unit boundaries and ecological features were accessed from DIVA-GIS (https://www.diva-gis.org/gdata). These were combined with data from home-based HIV testing, to conduct geospatial analysis of new HIV diagnoses. For spatial analysis, client data were aggregated to the sub-location where they were tested for HIV, and sub-location-level

geographic units were used for analysis and mapping. Village-level analysis was not possible owing to small numbers and lack of household-level point coordinates.

### *Global Moran's I statistic*

The Global Moran's I statistic was computed using GeoDa software tool version 1.12.1.131 (Anselin, Ibnu, & Youngihn, 2016; Anselin, Syabri, & Kho, 2010) in order to assess the presence of spatial autocorrelation of new HIV diagnoses at a sub-location level. A significant positive autocorrelation indicates the existence of either high-value or low-value clustering, while a negative autocorrelation indicates a tendency toward the juxtaposition of high values next to low values.

### *Kulldorff's spatial scan statistic*

The Kulldorff's spatial scan statistic (Kulldorff & Nagarwalla, 1995) was implemented using SaTScan™ version 9.6 (Martin Kulldorff and Information Management Services Inc, 2009) to detect spatial clusters of new HIV diagnoses. Since the proportion of clients newly diagnosed HIV positive was low, a discrete Poisson probability model was used for scanning. SaTScan™ software cyclically scans a window across space, calculating the number of observed and expected cases inside the window at each location, and adjusting for spatial in homogeneity of the background population.

The window with the maximum likelihood estimate is considered to be the most likely cluster, rejecting the null hypothesis of no clusters at p value $<0.05$. For this study, the Kulldorff spatial cluster detection looped over all the 161 sub-locations included in the analysis. A maximum spatial cluster size radius of five kilometers was used to inform HIV program implementation meaningfully at a granular sub-location level. Because Siaya County has a generalized epidemic,

and it was not possible to segregate the population proportion at higher risk, it was assumed that 50% of the total population were at risk of HIV-infection (excluding people living with HIV with previously known HIV status) (Martin Kulldorff and Information Management Services Inc, 2018). The maximum number of standard Monte Carlo replications was set to 999. Significant clusters were reported together with corresponding radii, number of observed and expected cases, relative risk, likelihood ratio and p-values. Clusters with a relative risk of >1.0 at p value <0.05 were considered significant clusters of higher new HIV diagnoses, while those with a relative risk of <1.0 at p value <0.05 were considered significant clusters of lower new HIV diagnoses. A standard Geographical Information System (GIS) program, Quantum GIS version 3.6 (QGIS.org), was used to map clusters and layer them over ecological features.

***Bayesian hierarchical spatial model***

A Bayesian hierarchical spatial model was used to assess the relationship between new HIV diagnosis and covariates while accounting for spatial autocorrelation in the data. A Bayesian estimation based on an Integrated Nested Laplace approximation (INLA) was computed using the R-INLA package (Blangiardo, Cameletti, Baio, & Rue, 2013; R Core Team, 2013). In a Bayesian framework, random effects are unknown quantities assigned to prior distributions that reflect prior knowledge on the structure of the effects, while enabling accounting for heterogeneity across spatial units. A Bayesian approach was applied to client-level and spatial parameters, separately and jointly.

The outcome of the analysis was new HIV positive diagnosis. The covariates: age, sex, marital status, time since last HIV test and sub-location proportion of total HIV positive clients, were included in the Bayesian spatially model.

$Y_{ijklm}$ denoted the number of new HIV positive individuals diagnosed among the $n_{ijklm}$ tested for HIV in the $i$-th sub-location for the $j$-th age category, $k$-th sex, $l$-th marital status and$m$-th time since last HIV test. It was assumed that $Y_{ijklm}$ is a Poisson random variable with mean $E_{ijklm}\theta_{ijklm}$. That is, $Y_{ijklm} \sim Poisson(E_{ijklm}\theta_{ijklm})$, where $E_{ijklm}$ denotes the expected number of cases and $\theta_{ijklm}$ is the "true" but unknown relative risk in the $i$-th sub-location for the $j$-th age category, $k$-th sex, $l$-th marital status and $m$-th time since last HIV test.

The Besag-York-Mollié (BYM) model (Besag, York, & Mollié, 1991; Blangiardo et al., 2013) was used, of the form:

$$\log(\pi_{ijklm}) = \beta_0 + \boldsymbol{X}_{ijklm}\boldsymbol{\beta} + u_i + v_i$$

where $\beta_0$ is the intercept that represents the overall log-odds of a new HIV positive diagnosis; $\boldsymbol{\beta}$ is a vector of parameters associated with the vector of covariate $\boldsymbol{X}_{ijklm}$; $u_i$ is a spatial structured component modeled with a conditional autoregressive (CAR) distribution $u_i|\boldsymbol{u}_{-i} \sim N\left(\bar{u}_{\delta_i}, \frac{\sigma_u^2}{n_{\delta_i}}\right)$, where $\bar{u}_{\delta_i} = n_{\delta_i}{}^{-1} \sum_{j \in \delta_i} u_j$, $\delta_i$ and $n_{\delta_i}$ represent the set of neighbors and the number of neighbors of sublocation *i* respectively; and $v_i$ is an unstructured spatial effect defined as $v_i \sim N(0, \sigma_v^2)$. The Besag York Mollié Poisson model (Besag, York, & Mollié, 1991) includes an ordinary random-effects component for non-spatial heterogeneity.

The posterior distributions of the parameters in the Bayesian spatial model were estimated via an Integrated Nested Laplace Approximation (INLA) approach in R statistical package, borrowing strength across sub locations to produce smoothed sub location level estimates even where the data were sparse. Full list of the latent models, likelihoods and prior

assumptions can be found in the R-INLA website at http://www.r-inla.org/ ("The R-INLA project,," ; Rue, Martino, & Chopin, 2009).

Unadjusted relative risk (uRR) and 95% Bayesian credible intervals (CIs) were computed to describe univariate associations. A multivariable Bayesian spatial Poisson model was used to assess the performance of four non-spatial and spatial models: fixed effects only, fixed effects in a spatially unstructured model, fixed effects in a spatially structured model, and fixed effects in a convolution unstructured and structured spatial random effects model. The convolution model, additionally allows for both spatially structured and unstructured heterogeneity in one model (Mollié, 1996). Measures of adjusted relative risk (aRR), 95% Bayesian CIs, precision of the spatially unstructured and structured random effect model, and the deviance information criterion (smaller values indicating better model performance) were reported.

Random effects maps of residual variability of new HIV diagnoses, not accounted for by the explanatory variables, were generated from the convolution Bayesian Poisson model, and mapped using ggplot2 R package (Wickham, 2016). These included unstructured random effects maps, showing variability when spatial autocorrelation was not taken into account, and structured random effects maps, when spatial autocorrelation was accounted for.

### 3.4 Methods for study objective 3: The use of mapping HIV testing uptake in identifying areas with low testing uptake yet higher HIV positive yield

This section describes the methods for study objective 3, under the following sub-headings:

- Study design, area, procedures for home-based HIV testing, study population, sample size and data management
- Data analysis

### 3.4.1 Study design, area, procedures for home-based HIV testing, study population, sample size and data management

The study design, area, procedures for home-based HIV testing, study population, sample size and data management for Study Objective 3 were similar to those described in Study Objective 2 above.

### 3.4.2 Data analysis

The socio-demographic characteristics of all individuals enumerated and tested for HIV were summarized using frequencies, percentages, medians, and interquartile ranges. Bivariable and multivariable analyses were conducted to assess the association between socio-demographic characteristics and HIV testing uptake.

The reasons for not testing among eligible clients were assessed, further describing clients who were not found at home and those who declined HIV testing. Chi-square tests were used to assess the difference between clients who tested for HIV and those not found at home or who declined testing.

Spatial data for administrative unit boundaries and ecological features were accessed from DIVA-GIS (https://www.diva-gis.org/gdata). These were combined with data from home-based HIV testing, to map HIV testing uptake. Mapping of granular sub-location levels of HIV testing uptake was done, identifying areas with high and low testing uptake. Quantiles of testing uptake were mapped and overlaid on sub-location clusters of new HIV diagnoses. The quantile proportion of clients who were not found at home and those who declined testing were also mapped and overlaid on sub-location clusters of new HIV diagnoses.

A standard Geographical Information System (GIS) program, Quantum GIS version 3.6 (QGIS.org), was used to map HIV testing uptake, the proportion of clients who were not found at

home and those who declined testing, and overlay on sub-location clusters of new HIV diagnoses.

## 3.5 Ethical considerations

### 3.5.1 Institutional Ethics Review

The Institutional Review Boards of Maseno University (Maseno, Kenya) and Kenyatta National Hospital (Nairobi, Kenya) approved the protocol to conduct this analysis (Appendices 6 and 7). The protocol was also reviewed in accordance with the United States Centers for Disease Control and Prevention, Atlanta, Georgia, (CDC) human research protection procedures and was determined to be research, but CDC staff did not interact with or have access to identifiable data or specimens for research purposes (Appendices 8 and 9).

### 3.5.2 Potential Risks and Benefits

There was minimal risk in participating in this study given that the study relied on data collected during the routine provision of HIV testing services. The potential minimal risk was inadvertent access of identifiable participant information by staff not authorized or breach of data confidentiality. To address this, routine data were collected, stored, and analyzed following strict security procedures and following Kenya Ministry of Health data handling guidelines. All data included in the analysis were de-identified, including the removal of patient names, unique numbers and any other identifiable information, and assigned a unique study-specific identifiers that could not be linked back to individual patient records. All findings were reported in aggregate form. Safeguards for protecting confidentially of data were strictly enforced. All staff who participated in the study had undergone trainings on human subjects' protections. The results of this study did not directly benefit individual clients. However, the results facilitated better understanding of HIV testing, and informed improvement of services for the targeted

populations.

### 3.5.3 Consent for HIV Testing and HIV Notification

The Kenya AIDS Indicator Survey, 2012, showed that by 15 years of age, 11.6% of adolescent girls and 20.2% of adolescent boys already had sex at least once in their lifetime (National AIDS STI Control Programme Ministry of Health Kenya, 2013). Based on this, the 2015 Kenya HIV testing guidelines recommended that adolescents and youth of 15 years and above can give their own consent for testing without the parent/guardian's consent (Kenya Ministry of Health, 2015). As per Kenya Ministry of Health HIV testing guidelines (Kenya Ministry of Health, 2015), testing for persons≥15 years in programmatic settings includes all the following: pretest counseling, verbal consent, testing, provision of test results, post-test counseling, and counseling on HIV prevention. Pre-test counseling, test results and post-test counseling are provided to the client immediately during the testing encounter. Since this study used data collected during the routine provision of HIV testing services, no additional consent was sought for study participation.

### 3.5.4 Request for Waiver of Informed Consent

Informed consent from study participants was not sought. The study requested for the waiver of informed consent. The request for waiver is based on the following reasons as indicated in Electronic Code of Federal Regulations 45CFR 46.116: [https://www.ecfr.gov/cgi-bin/text-idx?SID=bab035adde8f40a28cad9bc7f90232b4&mc=true&node=sp21.1.50.b&rgn=div6:

- There was no more than minimal risk to participants involved in the study.
- The study involved no procedures for which written consent is normally required outside of the public health programmatic context.

- The data used for this study were collected during the provision of routine health care services; and it was practically difficult to get back clients to the facilities for consenting.
- The data used for analysis in this study did not include identifiable information.
- The waiver or alteration did not adversely affect the rights and welfare of the participants; and the study results have been shared widely.

### 3.5.5 Incidents and adverse events reporting and management

All staff were aware that any incidents or adverse events were to be reported immediately to the US Centers for Disease Control and Prevention- Associate Director for Science office, and to the local Kenya Institutional Review Boards (Kenyatta National Hospital and Maseno University institutional review boards). There were no incidents or adverse events that occurred in the course of this study.

# CHAPTER FOUR
# RESULTS

## 4.1 Introduction to results

The study results show that the following strategies identified sub-populations and granular-geographic areas with higher HIV positive yield that should be targeted in the implementation of HIV testing services: a HIV predictive risk-score screening algorithm that identified patients with higher HIV-prevalence; geospatial analysis of new HIV diagnoses that identified sub-location clusters (<5 kilometers) of new HIV diagnoses; and mapping of HIV testing uptake that identified granular-geographic areas with low HIV testing uptake yet higher HIV positive yield.

The results for the study's three main objectives are described below.

## 4.2 Results for study objective 1: The use of a HIV predictive risk-score screening algorithm

### 4.2.1 Characteristics of patients at the five health facilities used for HIV predictive risk-score algorithm development

Out of the 45 total months (9-months for each of the 5 health facilities) that data were eligible for inclusion in the study, data for 37 (82%) months met the inclusion criteria. During these months, 99.9% (27,685/27,692) of persons ≥15 years attending outpatient services were screened for HIV testing eligibility, and 87% (21,764/24,966) of those eligible were tested for HIV. Of 21,745 patients with positive or negative HIV test results, 19,458 (89%) had behavioral risk characteristics documented and were included in the study analysis.

Among the 19,458 patient records included, the median age was 29 years (interquartile range, 22–43 years) and 11,149 (57%) were female [Table 4.1(a)]. Most patients [10,731 (61%)] were in monogamous marriage, and approximately two-thirds were either in trade/sales/service occupation [5,467 (29%)] or were school/college going [5,167 (27%)]. Most patients [18,450

(95%)] reported having sex in the prior 12 months, of whom 5,038 (28%) reported having two or more sexual partners, and 2,749 (17%) reported changes in sexual partners [Table 4.1(b)]. Among those with changes in sexual partners, 1,411 (51%) reported new sexual partners and 800 (29%) were widowed. Few patients reported having sex in exchange for money/favors/gifts [773 (4%)], having sex under the influence of alcohol/other substances [496 (3%)], having been coerced to have sex [480 (3%)], or having received treatment for STI in the prior 12 months [251 (1%)]. A minority of patients had never been tested for HIV [688 (3%)] or had a negative HIV test result >12 months prior [12 (0.1%)] (Table 4.1a). Overall, 210 (1.1%) patients were HIV positive.

Compared to female, a significantly higher proportion of males were never married (30% vs 23%, p <0.001), in a polygamous marriage (9% vs 3%, p <0.001), in a manual/domestic occupation (12% vs 1%, p <0.001), had $\geq$2 sexual partners (35% vs 22%, p <0.001) and reported a new sexual partner in the prior 12 months (13% vs 6%, p <0.001). Conversely, a significantly higher proportion of females were in monogamous marriage (64% vs 57%, p <0.001), widowed (7% vs 2%, p 0.004), in a trade/sale/service occupation (32% vs 24%, p <0.001), unemployed (19% vs 11%, p <0.001), or reported being widowed in the prior 12 months (6% vs 2%, p 0.051), (Tables 4.1a and 4.1b).

**Table 4.1 (a): Sociodemographic and HIV testing characteristics of outpatient attendees by gender at five high-volume facilities[a] used for algorithm development**

| Characteristic | All patients | Men | Women | p value |
|---|---|---|---|---|
| | n (%) | n (%) | n (%) | |
| Total | 19,458 | 8,309 | 11,149 | |
| **Sociodemographic characteristics** | | | | |
| Age in years, median (interquartile range) | 29 (22–43) | 30 (22–43) | 29 (22–45) | |
| Age categories | | | | |
| 15–19, 20–24 and ≥50 years | 10,577 (54%) | 4,399 (53%) | 6,178 (55%) | **0.042** |
| 25–29, 30–34 and 45–49 years | 6,211 (32%) | 2,668 (32%) | 3,543 (32%) | 1.00 |
| 35–39 and 40–44 years | 2,670 (14%) | 1,242 (15%) | 1,428 (13%) | 0.14 |
| Marital status | | | | |
| Never married | 4,546 (26%) | 2,248 (30%) | 2,298 (23%) | **<0.001** |
| Married monogamous | 10,731 (61%) | 4,312 (57%) | 6,419 (64%) | **<0.001** |
| Married polygamous | 952 (6%) | 667 (9%) | 285 (3%) | **<0.001** |
| Cohabiting | 239 (1%) | 59 (1%) | 180 (2%) | 0.61 |
| Separated/divorced | 224 (1%) | 118 (1%) | 106 (1%) | 0.54 |
| Widowed | 948 (5%) | 181 (2%) | 767 (7%) | **0.004** |
| Occupation | | | | |
| Professional/administrative/clerical | 2,180 (11%) | 1,011 (12%) | 1,169 (11%) | 0.15 |
| Manual[b]/domestic | 1,044 (6%) | 931 (12%) | 113 (1%) | **<0.001** |
| Agriculture | 2,285 (12%) | 1,050 (13%) | 1,235 (11%) | 0.14 |
| Trade/sales/service | 5,467 (29%) | 1,926 (24%) | 3,541 (32%) | **<0.001** |
| Unemployed | 2,924 (15%) | 863 (11%) | 2,061 (19%) | **<0.001** |
| School/college going | 5,167 (27%) | 2,282 (28%) | 2,885 (26%) | 0.11 |
| **HIV testing information** | | | | |
| Reason for HIV testing eligibility | | | | |
| Never been tested for HIV | 688 (3%) | 358 (4%) | 330 (3%) | 0.48 |
| HIV negative test >12 months prior | 12 (0.1%) | 7 (0.08%) | 5 (0.05%) | 0.84 |
| HIV negative test 6 to 12 months prior | 5,967 (31%) | 2,651 (32%) | 3,316 (30%) | 0.10 |
| HIV negative test 3 to 6 months prior | 9,454 (49%) | 3,913 (47%) | 5,541 (50%) | **0.004** |
| HIV negative test <3 months ago (unverified) | 2,666 (14%) | 1,067 (13%) | 1,599 (14%) | 0.46 |
| HIV negative test date unknown | 668 (3%) | 311 (4%) | 357 (3%) | 0.48 |
| Has tuberculosis, STI or recent HIV exposure | 3 (0.02%) | 2 (0.02%) | 1 (0.009%) | 0.95 |

[a]Jaramogi Oginga Odinga Teaching and Referral Hospital, Homa Bay County Hospital, Siaya County Hospital, Ahero Sub- County Hospital, Mbita Sub- County Hospital.
[b]Manual occupation refers to both skilled and unskilled.
P value significant at 0.05.
Abbreviations: n, number; STI, sexually transmitted infection.

**Table 4.1 (b): Behavioral characteristics of outpatient attendees by gender at five high-volume facilities[a] used for algorithm development**

| Characteristic | All patients | Male | Female | p value |
|---|---|---|---|---|
| | n (%) | n (%) | n (%) | |
| Total | 19,458 | 8,309 | 11,149 | |
| Had sex in the prior 12 months | | | | |
| Yes | 18,450 (95%) | 7,840 (94%) | 10,610 (95%) | **0.003** |
| No | 1,008 (5%) | 469 (6%) | 539 (5%) | 0.49 |
| Number of sexual partners in the prior 12 months | | | | |
| 1 | 13,220 (72%) | 5,084 (65%) | 8,136 (78%) | **<0.001** |
| ≥2 | 5,038 (28%) | 2,681 (35%) | 2,357 (22%) | **<0.001** |
| Changes in sexual partners in the prior 12 months | | | | |
| Not had a change in sexual partner | 13,523 (83%) | 5,459 (81%) | 8,064 (85%) | **<0.001** |
| New sexual partner | 1,411 (8%) | 867 (13%) | 544 (6%) | **<0.001** |
| Newly married | 155 (1%) | 69 (1%) | 86 (1%) | 1.00 |
| Ended a sexual relationship | 293 (2%) | 131 (2%) | 162 (2%) | 1.00 |
| Divorced/separated | 90 (1%) | 44 (1%) | 46 (0.4%) | 0.49 |
| Widowed | 800 (5%) | 175 (2%) | 625 (6%) | 0.051 |
| Had sex in exchange for money/favors in the prior 12 months[b] | n=17,373 | n=7,358 | n=10,015 | |
| Yes | 773 (4%) | 305 (4%) | 468 (5%) | 0.52 |
| Had sex under the influence of alcohol/other substances in the prior 12 months | n=17,366 | n=7,354 | n=10,012 | |
| Yes | 496 (3%) | 321 (4%) | 175 (2%) | 0.23 |
| Coerced to have sex in the prior 12 months | n=17,094 | n=7,274 | n=9,820 | |
| Yes | 480 (3%) | 97 (1%) | 383 (4%) | 0.15 |
| Reported treatment for STI in the prior 12 months | n=16,928 | n=7,188 | n=9,740 | |
| Yes | 251 (1%) | 121 (2%) | 130 (1%) | 0.51 |
| Engaged in sex work, men who have with men, female anal sex, injecting drugs for pleasure in the prior 12 months[b] | n=16,450 | n=6,960 | n=9,490 | |
| Yes | 730 (4%) | 294 (4%) | 436 (5%) | 0.53 |
| Circumcision status (males only) | | n=6,158 | | |
| Circumcised | | 4871 (78%) | | |

[a]Jaramogi Oginga Odinga Teaching and Referral Hospital, Homa Bay County Hospital, Siaya County Hospital, Ahero Sub- County Hospital, Mbita Sub- County Hospital.

[b]Due to multicollinearity, the characteristic "engaged in sex work, men who have with men, female anal sex, injecting drugs for pleasure in the prior 12 months" was excluded in univariable and multivariable analysis, while "had sex in exchange for money/favors in the prior 12 months" was included.

P value significant at 0.05.

Abbreviations: n, number; STI, sexually transmitted infection.

### 4.2.2 Overall HIV predictive risk-score algorithm development

The following characteristics were positively significantly associated with HIV infection in univariable analysis: being aged 35–39 and 40–44 years; male gender; manual/domestic and trade/sales/service occupation; polygamous marriage, separated/divorced or widowed; in the

prior 12 months having a new sexual partner, ≥2 sexual partners, or reporting treatment for STI; having never been tested for HIV; or having a negative HIV test result >12 months prior (Table 4.2).

**Table 4.2: Univariable association of socio-demographic and behavioral characteristics with HIV infection at five high-volume facilities used for algorithm development**

| Characteristic | Number HIV positive/Tested (%) | Univariable analysis[a] | |
|---|---|---|---|
| | | Odds ratio (95% CI) | p value |
| Total | 210/19,458 (1.08) | | |
| Ages 15–19, 20–24 and ≥50 years | 74/10,577 (0.70) | 0.46 (0.32, 0.64) | **<0.001** |
| Ages 25–29, 30–34 and 45–49 years | 79/6,211 (1.27) | 1.2 (0.87, 1.66) | 0.28 |
| Ages 35–39 and 40–44 years | 57/2,670 (2.13) | 2.39 (1.68, 3.39) | **<0.001** |
| Male | 110/8,309 (1.32) | 1.46 (1.06, 2.01) | **0.019** |
| Female | 100/11,149 (0.90) | 0.68 (0.5, 0.94) | **0.019** |
| Never married | 32/4,546 (0.70) | 0.5 (0.31, 0.83) | **0.006** |
| Married monogamous | 112/10,731 (1.04) | 0.74 (0.53, 1.02) | **0.06** |
| Married polygamous | 20/952 (2.10) | 2.11 (1.25, 3.56) | **0.005** |
| Cohabiting | 2/239 (0.84) | 0.88 (0.22, 3.58) | 0.86 |
| Separated/divorced | 16/224 (7.14) | 7.56 (4.01, 14.25) | **<0.001** |
| Widowed | 17/948 (1.79) | 2.3 (1.32, 4.02) | **0.003** |
| Professional/administrative/clerical occupation | 22/2,180 (1.01) | 0.71 (0.42, 1.22) | 0.22 |
| Manual/domestic occupation | 22/1,044 (2.11) | 2.06 (1.26, 3.39) | **0.004** |
| Trade/sales/service occupation | 93/5,467 (1.70) | 1.84 (1.33, 2.53) | **<0.001** |
| Agriculture occupation | 22/2,285 (0.96) | 0.82 (0.49, 1.39) | 0.47 |
| School/college going | 18/5,167 (0.35) | 0.24 (0.13, 0.45) | **<0.001** |
| Unemployed | 28/2,924 (0.96) | 1.22 (0.77, 1.95) | 0.39 |
| ≥2 sexual partners in the prior 12 months | 83/5,038 (1.65) | 2.29 (1.62, 3.22) | **<0.001** |
| No change in sexual partners in prior 12 months | 126/13,523 (0.93) | 0.52 (0.36, 0.77) | **0.001** |
| New sexual partner in the prior 12 months | 27/1,411 (1.91) | 2.61 (1.69, 4.03) | **<0.001** |
| Newly married in the prior 12 months | 1/155 (0.65) | 0.66 (0.09, 4.78) | 0.68 |
| Ended a sexual relationship in the prior 12 months | 2/293 (0.68) | 0 (0, 1.01) | 0.97 |
| Divorced/separated in the prior 12 months | 4/90 (4.44) | 3.09 (0.75, 12.76) | 0.12 |
| Widowed in the prior 12 months | 6/800 (0.75) | 1.03 (0.45, 2.34) | 0.94 |
| Had sex in exchange for money/favors in prior 12 months | 13/773 (1.68) | 1.58 (0.83, 3.02) | 0.16 |
| Had sex under influence of alcohol/other substance in the prior 12 months | 5/496 (1.01) | 0.77 (0.24, 2.42) | 0.65 |
| Coerced to have sex in the prior 12 months | 8/480 (1.67) | 1.43 (0.63, 3.26) | 0.39 |
| Reported treatment for STI in the prior 12 months | 9/251 (3.59) | 3.34 (1.55, 7.22) | **0.002** |
| Never been tested for HIV | 32/688 (4.65) | 5.44 (3.4, 8.71) | **<0.001** |
| HIV negative test >12 months prior | 1/12 (8.33) | 13.89 (1.7, 113.55) | **0.014** |
| HIV negative test 6 to 12 months prior | 63/5,967 (1.06) | 1.13 (0.8, 1.58) | 0.49 |
| HIV negative test 3 to 6 months prior | 74/9,454 (0.78) | 0.61 (0.44, 0.85) | **0.003** |
| HIV negative test <3 months ago (unverified) | 36/2,666 (1.35) | 0.93 (0.57, 1.53) | 0.79 |
| HIV negative test date unknown | 4/668 (0.60) | 0.58 (0.19, 1.84) | 0.36 |
| Has tuberculosis, STI or recent HIV exposure | 0/3 (0.00) | | |

[a]Missing data omitted from univariable analysis. P value significant at 0.05.
Abbreviations: n, number; CI, confidence interval; STI, sexually transmitted infection.

The initial full multivariable analysis included all the variables that were positively significantly associated with HIV infection in the univariable analysis. Additional variables that were also included based on known association with HIV infection were: divorced/separated or widowed, in the prior 12 months having sex in exchange for money/favors, and coerced to have sex (Table 4.3). The AUC for the full model was 0.66 (95% CI 0.44–0.88).

The final best-fit risk-score algorithm consisted of the following variables (Table 4.3), also shown in the scores for each variable:

- age category 35–39/40–44 years- **score 8**
- manual/domestic occupation- **score 8**
- trade/sales/service occupation- **score 7**
- polygamous marriage- **score 6**
- separated/divorced- **score 9**
- widowed- **score 17**
- in the prior 12 months having ≥2 sexual partners- **score 5**
- reporting treatment for an STI- **score 11**
- having never been tested for HIV- **score 18**
- having a negative HIV test result >12 months prior- **score 22**

For application at a health facility setting, a patient who has any of these characteristics, would be assigned a score as shown for each variable met. The total score for a patient would be derived by summing all the scores for variables met.

**Table 4.3: Multivariable association of socio-demographic and behavioral characteristics with HIV infection among outpatient attendees at five high-volume facilities**

| | Full multivariable model | | Stepwise multivariable analysis | | |
|---|---|---|---|---|---|
| **Characteristic** | **Odds ratio** | **β (95% CI)** | **Odds ratio** | **β (95% CI)** | **Risk score[a]** |
| Ages 35–39 and 40–44 years | 2.12 | 0.75 (0.38, 1.12) | 2.16 | 0.77 (0.4, 1.14) | 8 |
| Male | 1.16 | 0.15 (-0.21, 0.52) | | | |
| Manual/domestic occupation | 1.99 | 0.69 (0.11, 1.26) | 2.20 | 0.79 (0.23, 1.35) | 8 |
| Trade/sales/service occupation | 1.92 | 0.65 (0.29, 1) | 1.95 | 0.67 (0.31, 1.02) | 7 |
| Married polygamous | 1.55 | 0.44 (-0.15, 1.04) | 1.80 | 0.59 (0.01, 1.17) | 6 |
| Widowed | 3.90 | 1.36 (0.66, 2.06) | 2.39 | 0.87 (0.26, 1.48) | 9 |
| Separated/divorced | 6.96 | 1.94 (1.19, 2.68) | 5.26 | 1.66 (0.98, 2.34) | 17 |
| ≥2 sexual partners in prior 12 months | 1.63 | 0.49 (-0.04, 1.02) | 1.58 | 0.46 (0.06, 0.86) | 5 |
| New sexual partner in prior 12 months | 1.27 | 0.24 (-0.4, 0.88) | | | |
| Divorced/separated in prior 12 months | 0.33 | "-1.1 (-2.73, 0.52) | | | |
| Widowed in the prior 12 months | 0.39 | "-0.95 (-2.03, 0.13) | | | |
| Coerced to have sex in prior 12 months | 1.32 | 0.28 (-0.61, 1.17) | | | |
| Had sex in exchange for money/favors in prior 12 months | 0.90 | "-0.11 (-0.85, 0.63) | | | |
| Reported treatment for STI in prior 12 months | 2.61 | 0.96 (0.13, 1.79) | 2.97 | 1.09 (0.29, 1.9) | 11 |
| Never been tested for HIV | 6.23 | 1.83 (1.34, 2.33) | 6.17 | 1.82 (1.33, 2.31) | 18 |
| HIV negative result >12 months ago | 7.92 | 2.07 (-0.14, 4.29) | 9.03 | 2.2 (0.02, 4.37) | 22 |

[a]Computed by multiplying the beta regression coefficients by 10 and rounding to the nearest integer. Abbreviations: β, regression coefficient; CI, confidence interval; STI, sexually transmitted infection.

The variables in the final algorithm were each assigned a risk-score, and each patient's risk-score was calculated as the sum of risk-scores for variables met. Patients were grouped into the following four risk-score categories: ≤9 [HIV prevalence 0.6% (95% CI 0.46–0.75)], 10–15 [HIV prevalence 1.35% (95% CI 0.85–1.84)], 16–29 [HIV prevalence 2.65% (95% CI 1.8–3.51)], and ≥30 [HIV prevalence 15.15% (95% CI 9.03–21.27)] (Table 4.4). The three highest risk-score categories (score ≥ 10) accounted for 55% of HIV positive patients identified, yet represented just 24% of the total patients tested for HIV. Similarly, patients in the two highest risk-score categories (score ≥16) accounted for 37% of HIV positive patients identified, yet represented just 10% of the total patients tested for HIV.

**Table 4.4: Final algorithm risk-score categories for HIV predictive algorithm development dataset from five high volume facilities**

| Risk-score categories for algorithm development dataset | | | | |
|---|---|---|---|---|
| Risk-score category | Number HIV positive/Tested | HIV prevalence, % (95% CI) | % of total HIV positive | % of total tests |
| ≤9 | 68/11,289 | 0.6% (0.46, 0.75) | 45% | 76% |
| 10–15 | 28/2,076 | 1.35% (0.85, 1.84) | 18% | 14% |
| 16–29 | 36/1,357 | 2.65% (1.8, 3.51) | 24% | 9% |
| ≥30 | 20/132 | 15.15% (9.03, 21.27) | 13% | 1% |
| Total[a] | 152/14,854 | 1.02% | | |

[a]Patients with missing data omitted from the analysis
Abbreviations: CI, confidence interval.

### 4.2.3 Overall HIV predictive risk-score algorithm validation

The validation dataset consisted of 11,330 patient records, of which 174 (1.54%) were HIV positive. The socio-0demographic and behavioral characteristics of patients in the validation dataset are shown in Table 4.5. In comparison with the five facilities used for algorithm development, Kisumu County Hospital's population had the following differences (Table 4.5): a higher proportion of patients ages 25-29, 30-34 and 45-49 years (36% vs 32%, p <0.001), and a lower proportion of patients ages 15-19, 20-24 and ≥50 years (51% vs 54%, p <0.001); a higher proportion of patients never married (35% vs 26%, p <0.001), and lower proportion of those in married polygamous relationships (55% vs 61%, p <0.001), and those reporting ≥2 sexual partners in the prior 12 months (11% vs 28%, p <0.001). The facility had a higher proportion of patients in manual domestic (9% vs 6%, p <0.001), trade/sales/service occupations (34% vs 29%, p <0.001), and those unemployed (18% vs 15%, p <0.005); and a lower proportion of patients in agriculture (6% vs 12%, p <0.001), and those school/college going (23% vs 27%, p <0.001). Additionally, the facility had a higher proportion of patients who reported a HIV negative test <3 months ago (18% vs 14%, p <0.001); and a lower proportion of patients who reported a HIV negative test >12 months prior (0.02% vs 0.1%, p <0.001); reported HIV negative test 3 to 6

months prior (43% vs 49%, p <0.001), and reported had tuberculosis, STI or recent HIV exposure (0.01% vs 0.02%, p <0.001).

**Table 4.5: Comparison of characteristics of patients at the five health facilities used for algorithm development and one facility used for algorithm validation**

| Characteristic | Five health facilities[a] | One health facility[b] | p value |
|---|---|---|---|
| | n (%) | n (%) | |
| Total | 19,458 | 11,330 | |
| **Sociodemographic characteristics** | | | |
| Age in years, median (interquartile range) | 29 (22–43) | 27 (22–38) | |
| Age categories | | | |
| 15–19, 20–24 and ≥50 years | 10,577 (54%) | 5,744 (51%) | **<0.001** |
| 25–29, 30–34 and 45–49 years | 6,211 (32%) | 4,089 (36%) | **<0.001** |
| 35–39 and 40–44 years | 2,670 (14%) | 1,497 (13%) | 0.37 |
| Gender | | | |
| Male | 8,309 (43%) | 4,706 (42%) | 0.27 |
| Female | 11,149 (57%) | 6,624 (58%) | 0.19 |
| Marital status | | | |
| Never married | 4,546 (26%) | 3,968 (35%) | **<0.001** |
| Married monogamous | 10,731 (61%) | 6,191 (55%) | **<0.001** |
| Married polygamous | 952 (6%) | 465 (4%) | 0.4 |
| Cohabiting | 239 (1%) | 11 (0.1%) | 0.74 |
| Separated/divorced | 224 (1%) | 154 (1%) | 1 |
| Widowed | 948 (5%) | 532 (5%) | 1 |
| Occupation | | | |
| Professional/administrative/clerical | 2,180 (11%) | 1,086 (10%) | 0.38 |
| Manual/domestic | 1,044 (6%) | 1,000 (9%) | **<0.001** |
| Agriculture | 2,285 (12%) | 643 (6%) | **<0.001** |
| Trade/sales/service | 5,467 (29%) | 3,783 (34%) | **<0.001** |
| Unemployed | 2,924 (15%) | 1,974 (18%) | **0.005** |
| School/college going | 5,167 (27%) | 2,531 (23%) | **<0.001** |
| ≥2 sexual partners in the prior 12 months | 5,038 (28%) | 1,181 (11%) | **<0.001** |
| Reported treatment for STI in the prior 12 months | 251 (1%) | 74 (1%) | 1 |
| Reason for HIV testing eligibility | | | |
| Never been tested for HIV | 688 (4%) | 695 (6%) | 0.08 |
| HIV negative test >12 months prior | 12 (0.1%) | 2 (0.02%) | **<0.001** |
| HIV negative test 6 to 12 months prior | 5,967 (31%) | 3,623 (32%) | 0.31 |
| HIV negative test 3 to 6 months prior | 9,454 (49%) | 4,836 (43%) | **<0.001** |
| HIV negative test <3 months ago (unverified) | 2,666 (14%) | 2,074 (18%) | **<0.001** |
| HIV negative test date unknown | 668 (3%) | 99 (1%) | 0.26 |
| Has tuberculosis, STI or recent HIV exposure | 3 (0.02%) | 1 (0.01%) | **<0.001** |

[a]Jaramogi Oginga Odinga Teaching and Referral Hospital, Homa Bay County Hospital, Siaya County Hospital, Ahero Sub- County Hospital, Mbita Sub- County Hospital. [b]Kisumu County Hospital. P value significant at 0.05.

Abbreviations: n, number; STI, sexually transmitted infection.

When applied to the validation dataset, the final risk-score algorithm/model had an AUC of 0.69 (95% CI 0.60–0.77) and $R^2$ of 0.88. The risk score categories ≤9, 10–15, 16–29 and ≥30 had an increasing HIV prevalence of 0.97% (95% CI 0.76–1.18), 2.32% (95% CI 1.47–3.17), 3.69% (95% CI 2.62–4.76) and 6.76% (95% CI 1.04–12.48), respectively (Table 4.6). The three highest risk-score categories (score ≥ 10) accounted for 49% of HIV positive patients identified, but only 23% of the total patients tested for HIV. The two highest risk-score categories (score ≥16) accounted for 31% of HIV positive patients identified, but only 12% of the total patients tested for HIV.

**Table 4.6: Final algorithm risk-score categories for HIV predictive algorithm validation dataset from one large volume facility**

| **Risk-score categories for algorithm validation dataset** | | | | |
|---|---|---|---|---|
| Risk-score category | Number HIV positive/Tested | HIV prevalence, % (95% CI) | % of total HIV positive | % of total tests |
| ≤9 | 79/8,142 | 0.97% (0.76, 1.18) | 51% | 77% |
| 10–15 | 28/1,207 | 2.32% (1.47, 3.17) | 18% | 11% |
| 16–29 | 44/1,193 | 3.69% (2.62, 4.76) | 28% | 11% |
| ≥30 | "5/74 | 6.76% (1.04, 12.48) | 3% | 1% |
| Total[a] | 156/10,616 | 1.47% | | |

[a]Patients with missing data were omitted from the analysis
Abbreviations: CI, confidence interval.

### 4.2.4 Development of gender-specific HIV predictive risk-score algorithms

Characteristics that were positively significantly associated with HIV infection in univariable analysis (OR>1.0 at p≤0.05) among male and female are shown in Table 4.7 (a). and Table 4.7 (b). Full multivariable models for males and females are shown in Tables 4.8 and 4.9. The AUC for the full model was 0.75 (95% CI 0.65–0.85) among males and 0.68 (95% CI 0.56–0.8) among females.

**Table 4.7 (a): Univariable association of sociodemographic and HIV testing characteristics with HIV infection by gender, at five high-volume facilities used for algorithm development.**

| Characteristic | Male | | | Female | | |
|---|---|---|---|---|---|---|
| | Number HIV positive/ Tested | Univariable analysis[a] | | Number HIV positive/ Tested | Univariable analysis[a] | |
| | | Odds ratio (95% CI) | p value | | Odds ratio (95% CI) | p value |
| Total | 110/8,309 | | | 100/11,149 | | |
| Ages 15–19, 20–24 and ≥50 years | 36/4,399 | 0.48 (0.35, 0.67) | **<0.001** | 38/6,178 | 0.41 (0.31, 0.54) | **<0.001** |
| Ages 25–29, 30–34 and 45–49 years | 41/2,668 | 1.63 (1.19, 2.24) | **0.002** | 38/3,543 | 1.25 (0.95, 1.63) | 0.11 |
| Ages 35–39 and 40–44 years | 33/1,242 | 1.58 (1.07, 2.34) | **0.022** | 24/1,428 | 2.75 (2.06, 3.67) | **<0.001** |
| Professional/administrative/clerical occupation | 7/1,011 | 0.63 (0.34, 1.17) | 0.15 | 15/1,169 | 0.93 (0.6, 1.45) | 0.75 |
| Manual/domestic occupation | 21/931 | 2.02 (1.26, 3.25) | **0.004** | 1/113 | 2.48 (1.7, 3.63) | **<0.001** |
| Trade/sales/service occupation | 48/1,926 | 2.28 (1.66, 3.13) | **<0.001** | 45/3,541 | 1.9 (1.45, 2.49) | **<0.001** |
| Agriculture occupation | 15/1,050 | 0.65 (0.34, 1.23) | 0.18 | 7/1,235 | 0.84 (0.52, 1.37) | 0.49 |
| School/college going | 6/2,282 | 0.19 (0.1, 0.37) | **<0.001** | 12/2,885 | 0.25 (0.15, 0.42) | **<0.001** |
| Unemployed | 10/863 | 1.09 (0.72, 1.66) | 0.68 | 18/2,061 | 0.74 (0.49, 1.11) | 0.14 |
| Never married | 11/2,248 | 0.41 (0.27, 0.65) | **<0.001** | 21/2,298 | 0.55 (0.39, 0.78) | **<0.001** |
| Married monogamous | 64/4,312 | 0.93 (0.68, 1.28) | 0.67 | 48/6,419 | 0.94 (0.72, 1.23) | 0.65 |
| Married polygamous | 12/667 | 2.1 (1.24, 3.55) | **0.005** | 8/285 | 1.87 (1.15, 3.04) | **0.012** |
| Widowed | 5/181 | 1.12 (0.55, 2.3) | 0.76 | 12/767 | 1.77 (1.11, 2.82) | **0.016** |
| Separated/divorced[b] | 11/118 | 9.29 (5.74, 15.04) | **<0.001** | 5/106 | 6.46 (3.53, 11.82) | **<0.001** |
| Cohabiting[c] | 1/59 | | | 1/180 | 1.31 (0.32, 5.32) | 0.71 |
| Never been tested for HIV | 19/358 | 2.24 (1.31, 3.83) | **0.003** | 13/330 | 3.78 (2.59, 5.51) | **<0.001** |
| HIV negative test >12 months prior[c] | 0/7 | | | 1/5 | 16.12 (1.88, 138.55) | **0.011** |
| HIV negative test 6 to 12 months prior | 31/2,651 | 1.04 (0.74, 1.45) | 0.83 | 32/3,316 | 0.81 (0.6, 1.08) | 0.15 |
| HIV negative test 3 to 6 months prior | 39/3,913 | 0.69 (0.5, 0.96) | **0.025** | 35/5,541 | 0.7 (0.53, 0.92) | **0.010** |
| HIV negative test <3 months ago (unverified) | 18/1,067 | 1.28 (0.85, 1.9) | 0.23 | 18/1,599 | 1.16 (0.82, 1.64) | 0.41 |
| HIV negative test date unknown | 3/311 | 0.94 (0.35, 2.54) | 0.90 | 1/357 | 1.13 (0.5, 2.57) | 0.76 |
| Has tuberculosis/STI/recent HIV exposure[cd] | 0/2 | | | 0/1 | | |

[a]Missing data were omitted from univariable analysis. [b]Due to multicollinearity, the characteristic "divorced/separated in prior 12 months" was excluded in multivariable analysis, while the characteristic "separated/divorced marital status" was included. [c]Characteristics omitted in univariable analysis among males, and [d]females due to small numbers. P value significant at 0.05. Abbreviations: n, number; CI, confidence interval; STI, sexually transmitted infection.

**Table 4.7 (b): Univariable association of behavioral characteristics with HIV infection by gender, at five high-volume facilities used for algorithm development**

| Characteristic | Male | | | Female | | |
|---|---|---|---|---|---|---|
| | Number HIV positive/ Tested | Univariable analysis[a] | | Number HIV positive/ Tested | Univariable analysis[a] | |
| | | Odds ratio (95% CI) | p value | | Odds ratio (95% CI) | p value |
| Total | 110/8,309 | | | 100/11,149 | | |
| ≥2 sexual partners in prior 12 months | 55/2,681 | 1.93 (1.38, 2.68) | **<0.001** | 28/2,357 | 1.65 (1.23, 2.22) | **<0.001** |
| No change in sexual partner in prior 12 months | 59/5,459 | 0.48 (0.32, 0.72) | **<0.001** | 67/8,064 | 0.75 (0.51, 1.12) | 0.16 |
| New sexual partner in prior 12 months | 20/867 | 2.52 (1.57, 4.05) | **<0.001** | 7/544 | 1.74 (1.08, 2.8) | **0.023** |
| Newly married in prior 12 months[cd] | 1/69 | | | 0/86 | | |
| Ended a sexual relationship in prior 12 months[c] | 0/131 | | | 2/162 | 0.4 (0.06, 2.9) | 0.37 |
| Divorced/separated in prior 12 months[b] | 1/44 | 6.53 (2.59, 16.49) | **<0.001** | 3/46 | 2.91 (0.71, 12.03) | 0.14 |
| Widowed in prior 12 months | 3/175 | 0.4 (0.1, 1.64) | 0.20 | 3/625 | 0.87 (0.39, 1.98) | 0.74 |
| Had sex in exchange for money/favors in prior 12 months | 9/305 | 1.58 (0.77, 3.23) | 0.22 | 4/468 | 1.05 (0.47, 2.39) | 0.90 |
| Had sex under influence of alcohol/other substance in prior 12 months | 3/321 | 1.09 (0.4, 2.97) | 0.86 | 2/175 | 1.91 (0.93, 3.9) | 0.08 |
| Coerced to have sex in prior 12 months[c] | 0/97 | | | 8/383 | 1.54 (0.68, 3.49) | 0.30 |
| Reported treatment for STI in prior 12 months | 3/121 | 1.67 (0.52, 5.29) | 0.39 | 6/130 | 4.02 (2.03, 7.95) | **<0.001** |
| Circumcised (males only) | 46/4,871 | 2.64 (1.59, 4.39) | **<0.001** | | | |

[a]Missing data were omitted from univariable analysis.

[b]Due to multicollinearity, the characteristic "divorced/separated in prior 12 months" was excluded in multivariable analysis, while the characteristic "separated/divorced marital status" was included.

[c]Characteristics omitted in univariable analysis among males, and [d]females due to small numbers.

Abbreviations: n, number; CI, confidence interval; STI, sexually transmitted infection.

The final best-fit model/risk-score algorithm among males had an AUC of 0.76 (95% CI 0.56–0.96) and an $R^2$ of 0.69, and consisted of the following variables: age categories 25–29/30–34/45–49 years and 35–39/40–44 years; occupation (manual/domestic or trade/sales/service); marital status (separated/divorced or widowed); in the prior 12 months having ≥2 sexual partners or a new sexual partner; circumcised status; and having never been tested for HIV (Table 4.8).

The final risk-score algorithm among females had an AUC of 0.66 (95% CI 0.47–0.85) and an $R^2$ of 0.87, and consisted of the following variables: age category 35–39/40–44 years; trade/sales/service occupation; marital status (polygamous marriage, separated/divorced or widowed); in the prior 12 months having a new sexual partner or reporting treatment for an STI; and having never been tested for HIV or having a negative HIV test result >12 months prior (Table 4.9).

**Table 4.8: Multivariable association of sociodemographic and behavioral characteristics with HIV infection among male at the five high-volume facilities**

| Characteristic | Full multivariable model | | Stepwise multivariable analysis | | |
|---|---|---|---|---|---|
| | Odds ratio | β (95% CI) | Odds ratio | β (95% CI) | Risk score[a] |
| Ages 25–29, 30–34 and 45–49 years[b] | 1.82 | 0.6 (0.11, 1.08) | 1.80 | 0.59 (0.1, 1.07) | 6 |
| Ages 35–39 and 40–44 years[b] | 2.78 | 1.02 (0.51, 1.54) | 2.78 | 1.02 (0.51, 1.54) | 10 |
| Manual/domestic occupation[b] | 3.57 | 1.27 (0.77, 1.77) | 3.52 | 1.26 (0.76, 1.76) | 13 |
| Trade/sales/service occupation[b] | 2.63 | 0.97 (0.49, 1.44) | 2.66 | 0.98 (0.5, 1.46) | 10 |
| Married polygamous[b] | 1.34 | 0.29 (-0.32, 0.9) | | | |
| Widowed[c] | 4.57 | 1.52 (0.55, 2.49) | 4.40 | 1.48 (0.51, 2.45) | 15 |
| Separated/divorced[b] | 4.36 | 1.47 (0.71, 2.23) | 4.22 | 1.44 (0.69, 2.19) | 14 |
| ≥2 sexual partners in prior 12 months[b] | 1.49 | 0.4 (-0.09, 0.89) | 1.57 | 0.45 (-0.02, 0.92) | 5 |
| New sexual partner in prior 12 months[b] | 1.46 | 0.38 (-0.24, 0.99) | 1.60 | 0.47 (-0.12, 1.05) | 5 |
| Had sex in exchange for money/favors in prior 12 months[c] | 1.25 | 0.22 (-0.69, 1.13) | | | |
| Had sex under influence of alcohol/other substance in prior 12 months[c] | 0.82 | -0.2 (-1.15, 0.75) | | | |
| Circumcised (males only)[b] | 2.66 | 0.98 (0.62, 1.34) | 2.67 | 0.98 (0.62, 1.35) | 10 |
| Reported treatment for STI in prior 12 months[c] | 1.99 | 0.69 (-0.4, 1.78) | | | |
| Never been tested for HIV[b] | 4.62 | 1.53 (1, 2.07) | 4.56 | 1.52 (0.98, 2.05) | 15 |

[a]Computed by multiplying the beta regression coefficients by 10 and rounding to the nearest integer.
[b]Variables with significant association with HIV infection in univariable analysis, and included in multivariable analysis.
[c]Variables without significant association with HIV infection in univariable analysis, but included in multivariable analysis based on prior knowledge of association with HIV infection.
Abbreviations: β, regression coefficient; CI, confidence interval; STI, sexually transmitted infection.

**Table 4.9: Multivariable association of sociodemographic and behavioral characteristics with HIV infection among female at the five high-volume facilities**

| Characteristic | Full multivariable model | | Stepwise multivariable analysis | | |
|---|---|---|---|---|---|
| | Odds ratio | β (95% CI) | Odds ratio | β (95% CI) | Risk score[a] |
| Ages 35–39 and 40–44 years[b] | 1.71 | 0.54 (0.15, 0.92) | 1.71 | 0.54 (0.15, 0.92) | 5 |
| Manual/domestic occupation[b] | 1.02 | 0.02 (-1.42, 1.45) | | | |
| Trade/sales/service occupation[b] | 1.94 | 0.66 (0.34, 0.99) | 1.95 | 0.67 (0.34, 1) | 7 |
| Married polygamous[b] | 3.19 | 1.16 (0.59, 1.73) | 3.19 | 1.16 (0.59, 1.73) | 12 |
| Widowed[b] | 2.55 | 0.94 (0.38, 1.49) | 2.67 | 0.98 (0.48, 1.49) | 10 |
| Separated/divorced[b] | 9.61 | 2.26 (1.63, 2.9) | 10.56 | 2.36 (1.81, 2.91) | 24 |
| ≥2 sexual partners in prior 12 months[b] | 1.15 | 0.14 (-0.42, 0.7) | | | |
| New sexual partner in prior 12 months[b] | 3.21 | 1.17 (0.38, 1.96) | 2.76 | 1.02 (0.38, 1.65) | 10 |
| Divorced/separated in prior 12 months[c] | 1.31 | 0.27 (-0.91, 1.45) | | | |
| Coerced to have sex in prior 12 months[c] | 1.47 | 0.38 (-0.54, 1.3) | | | |
| Had sex in exchange for money/favors in prior 12 months[c] | 0.16 | -1.86 (-3.49, -0.22) | | | |
| Had sex under influence of alcohol/other substance in prior 12 months[c] | 1.41 | 0.34 (-1.2, 1.89) | | | |
| Reported treatment for STI in prior 12 months[b] | 2.75 | 1.01 (0.04, 1.98) | 2.26 | 0.82 (-0.15, 1.79) | 8 |
| Never been tested for HIV[b] | 2.89 | 1.06 (0.45, 1.67) | 2.91 | 1.07 (0.46, 1.68) | 11 |
| HIV negative test >12 months prior[b] | 10.93 | 2.39 (0.2, 4.58) | 10.92 | 2.39 (0.19, 4.59) | 24 |

[a]Computed by multiplying the beta regression coefficients by 10 and rounding to the nearest integer.
[b]Variables with significant association with HIV infection in univariable analysis, and included in multivariable analysis.
[c]Variables without significant association with HIV infection in univariable analysis, but included in multivariable analysis based on prior knowledge of association with HIV infection.
Abbreviations: β, regression coefficient; CI, confidence interval; STI, sexually transmitted infection.

Risk-score categories and corresponding HIV prevalence among males and females are shown in Table 4.10. Among males, the three highest risk-score categories (score ≥13) accounted for 86% of HIV positive patients identified, yet represented 50% of the total patients tested for HIV. Similarly, among females, the three highest risk-score categories (score ≥8) accounted for 51% of HIV positive patients identified, yet represented 23% of the total patients tested for HIV (Table 4.10).

**Table 4.10: Final algorithm risk-score categories for development dataset from five high volume facilities, stratified by gender**

| **Male** | | | | |
|---|---|---|---|---|
| **Risk-score categories for algorithm development dataset** | | | | |
| Risk-score category | Number HIV positive/Tested | HIV prevalence, % (95% CI) | % of total HIV positive | % of total tests |
| ≤12 | 9/2,490 | 0.36% (0.17, 0.69) | 14% | 50% |
| 13–26 | 25/2,044 | 1.22% (0.79, 1.8) | 40% | 41% |
| 27–39 | 19/434 | 4.38% (2.66, 6.75) | 30% | 8% |
| ≥40 | 10/33 | 30.3% (15.59, 48.71) | 16% | 1% |
| Total | 63/5,001 | 1.26% | | |
| **Female** | | | | |
| **Risk-score categories for algorithm development dataset** | | | | |
| Risk-score category | Number HIV positive/Tested | HIV prevalence, % (95% CI) | % of total HIV positive | % of total tests |
| ≤7 | 37/6,676 | 0.55% (0.39, 0.76) | 49% | 77% |
| 8–20 | 25/1,767 | 1.41% (0.92, 2.08) | 33% | 20% |
| 21–27 | 8/135 | 5.93% (2.59, 11.34) | 11% | 2% |
| ≥28 | 5/66 | 7.58% (2.51, 16.8) | 7% | 1% |
| Total | 75/8644 | 0.87% | | |

### 4.2.5 Validation of the gender-specific HIV predictive risk-score algorithms

The validation dataset comprised 4,706 (42%) males and 6,624 (58%) females. When applied to the validation dataset, the final algorithm/model had an AUC of 0.71 (95% CI 0.57–0.86) and an $R^2$ of 0.85 among males, and an AUC of 0.66 (95% CI 0.49–0.84) and an $R^2$ of 0.95 among females. The risk-score categories and corresponding HIV prevalence among males and females are shown in Table 4.11.

**Table 4.11: Final algorithm risk-score categories for validation dataset from one large volume facility, stratified by gender**

| **Male** | | | | |
|---|---|---|---|---|
| **Risk-score categories for algorithm validation dataset** | | | | |
| Risk-score category | Number HIV positive/Tested | HIV prevalence, % (95% CI) | % of total HIV positive | % of total tests |
| ≤12 | 8/1,765 | 0.45% (0.2, 0.89) | 13% | 43% |
| 13–26 | 35/1,919 | 1.82% (1.27, 2.53) | 55% | 47% |
| 27–39 | 15/379 | 3.96% (2.23, 6.44) | 24% | 9% |
| ≥40 | 5/35 | 14.29% (4.81, 30.26) | 8% | 1% |
| Total | 63/4,098 | 1.54% | | |
| **Female** | | | | |
| **Risk-score categories for algorithm validation dataset** | | | | |
| Risk-score category | Number HIV positive/Tested | HIV prevalence, % (95% CI) | % of total HIV positive | % of total tests |
| ≤7 | 42/4,774 | 0.88% (0.63, 1.19) | 47% | 76% |
| 8–20 | 30/1,302 | 2.3% (1.56, 3.27) | 34% | 21% |
| 21–27 | 10/118 | 8.47% (4.14, 15.03) | 11% | 2% |
| ≥28 | 7/61 | 11.48% (4.74, 22.22) | 8% | 1% |
| Total | 89/6,255 | 1.42% | | |

## 4.3 Results for study objective 2: The use of geospatial analysis of new HIV diagnoses in identifying areas with higher HIV positive yield

### 4.3.1 Characteristics of home-based HIV testing clients

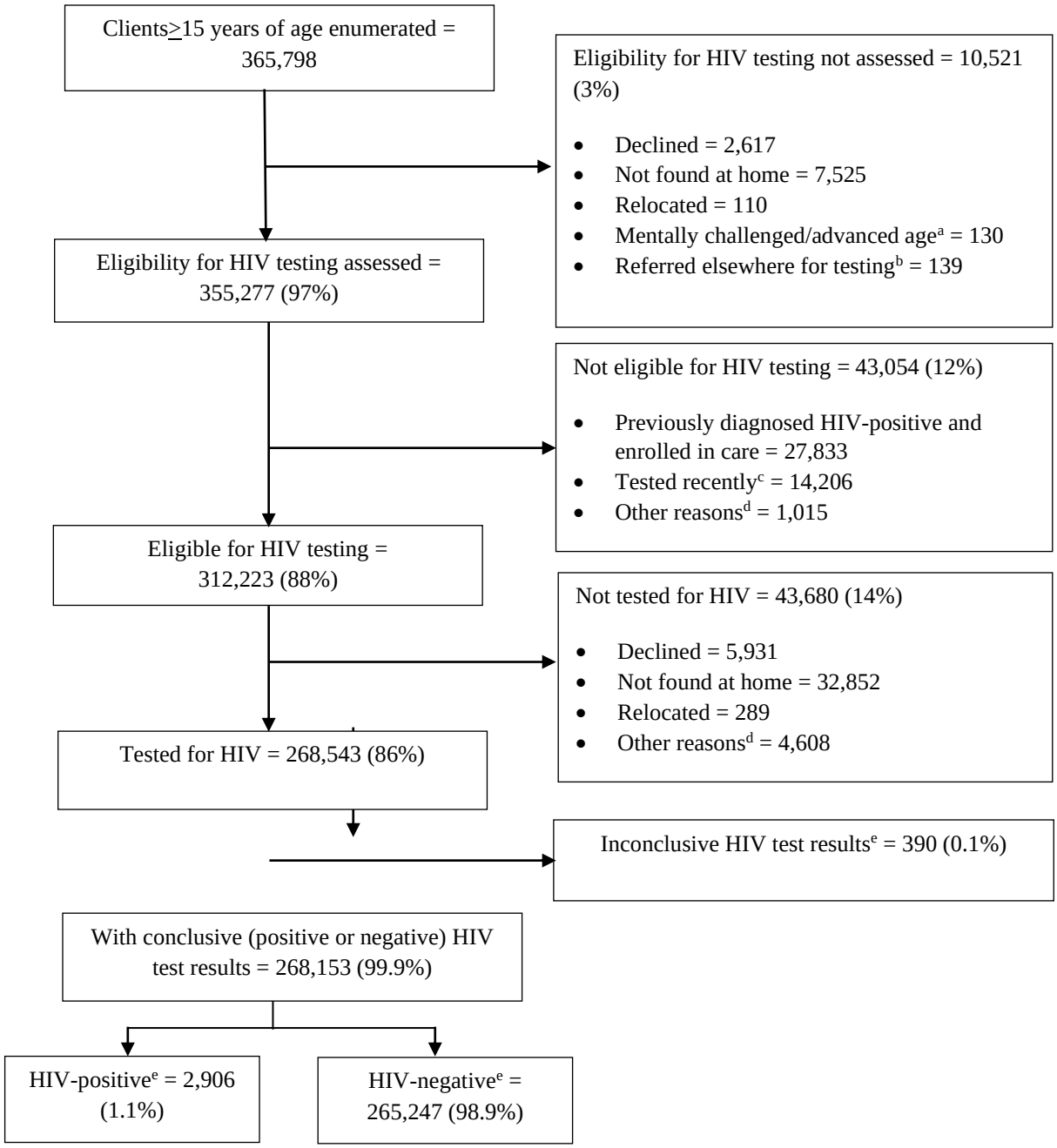
From the 161 Siaya administrative sub-locations included in the analysis, 365,798 clients aged ≥15 years from 136,607 households were enumerated for home-based HIV testing (Figure 4.1). Among those enumerated, 136,607 (37%) were household-heads, 80,161 (22%) were spouses, 110,255 (30%) were children aged ≥15 years, and 38,775 (11%) were other relatives/non-relatives (Table 4.12). Overall, those enumerated had a median age of 30 years (interquartile range 20–47 years), and 203,170 (56%) were females. Of the total clients enumerated, 355,277 (97%) were assessed for HIV testing eligibility, and 312,223 (88%) were eligible for testing (Figure 4.1, Table 4.12). Among those eligible, 268,543 (86%) were tested for HIV, and 2,906

(1.1%) of 268,153 clients with conclusive HIV test results were diagnosed HIV positive. Table 4.12 shows the characteristics of clients who received home-based HIV testing. Slightly more than a third (37% of those enumerated, 36% of those eligible and 35% of those tested) were household heads; about a third were children ≥15 years (30% of those enumerated, 32% of those eligible and 30% of those tested); and about a quarter were spouses (22% of those enumerated, 21% of those eligible and 23% of those tested). The median aged for those enumerated, eligible and tested for HIV was 30 (IQR 20-47) years, 28 (IQR 19-46) years and 28 (IQR 19-47) years, respectively. Majority were female (56% of those enumerate, 55% of those eligible, and 57% of those tested). About half (49%) of those tested were in married monogamous relationships, and 38% were single. Majority, 69%, had been tested for HIV 3-12 months prior. Table 4.12 and Figure 4.2 show the proportion of clients who tested HIV positive (HIV positive yield) during home-based testing. The highest HIV positive yield was among those 25-35 years of age (1.9% HIV positive yield); females (1.2% HIV positive yield); those separated/divorced (5.2% HIV positive yield) and married polygamous (2.5% HIV positive yield); and those tested for HIV >12 months prior (1.5% yield) or never tested (1.2% yield). The reasons for not testing among eligible clients are shown in Figure 4.1.

### 4.3.2 Sub-location distribution of different characteristics

The 161 sub-locations had a median HIV testing uptake among eligible clients of 87% (interquartile range 82%–91%), a median new HIV positive yield of 1.1% (interquartile range 0.8%–1.5%), and a median proportion of total HIV positive clients of 9.1% (interquartile range 7.6%–10.4%), (Table 4.12). Figure 4.3 shows the mapping of the proportion of different client characteristics at sub-location level, computed by the number of clients with the characteristic divided by the total number of clients in the sub-location. The maps show: a) The proportion of

clients tested HIV positive during home-based testing; b) The proportion of total HIV positive clients; c) The proportion of clients age >25 years; d) The proportion of female; e) The proportion married polygamous; f) The proportion separated/divorced; g) The proportion widowed; h) The proportion never tested for HIV; and i) The proportion tested for HIV >12 months ago.



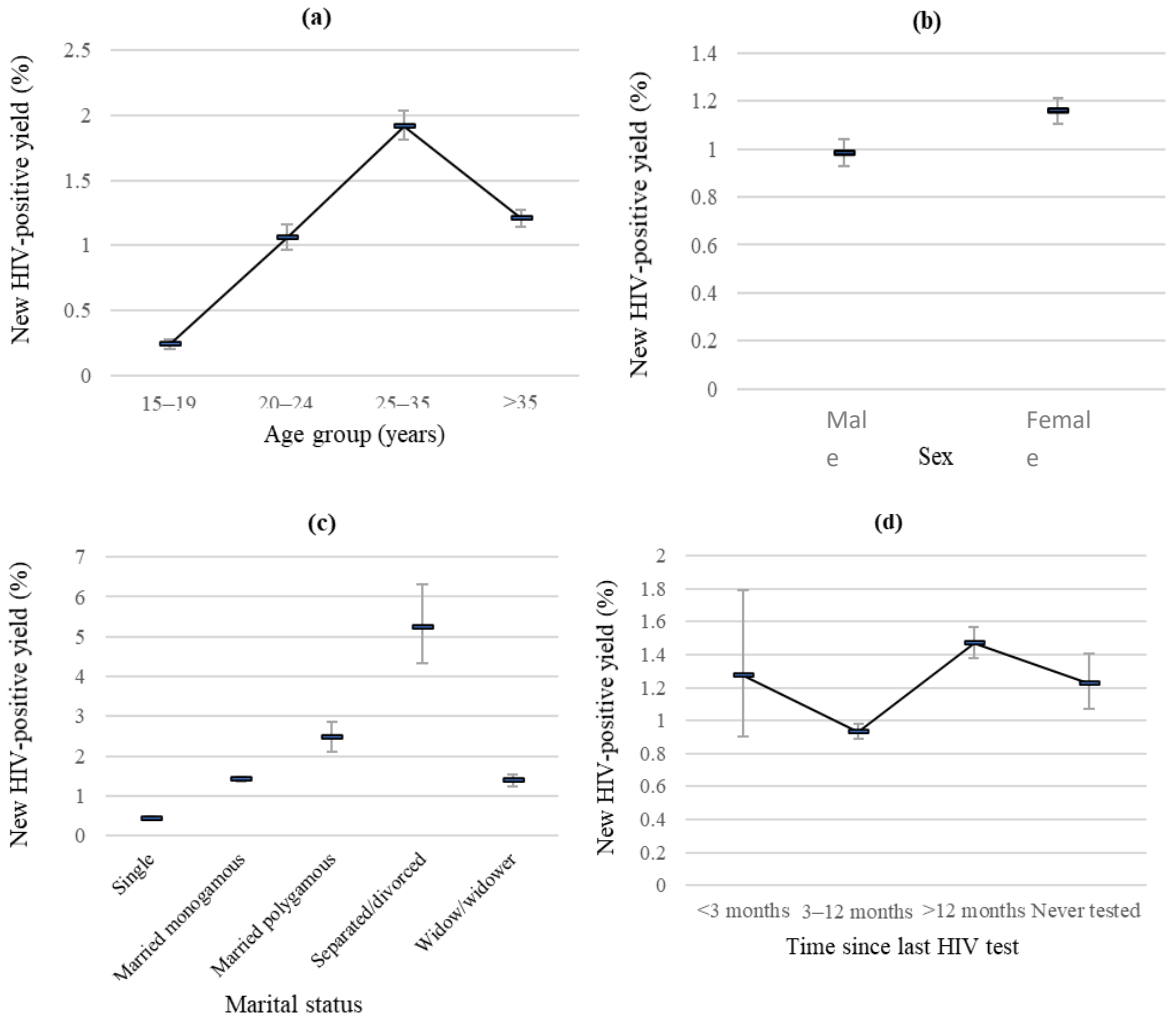

Note: [a]Advanced age referred to elderly clients who were unable to comprehend HIV testing due to their diminished mental capacity related to old age. [b]Clients 15–24 years of age in selected sub-locations were referred to another program offering testing for young people. [c]Self-reported tested recently within the prior 3 months. [d]Details of other reasons not given. [e]An individual was considered HIV-negative (uninfected) if the Determine test result was negative (considered a conclusive negative result), HIV-positive (infected) if both the Determine and First Response serial tests results were positive (considered a conclusive positive result), and inconclusive if the Determine test was positive and First Response test was negative.

**Figure 4.1: Flow diagram of clients receiving home-based HIV testing in Siaya County, May 2016 to July 2017**

**Table 4.12: Characteristics of clients aged ≥15 years offered home-based HIV testing in Siaya County, May 2016 to July 2017**

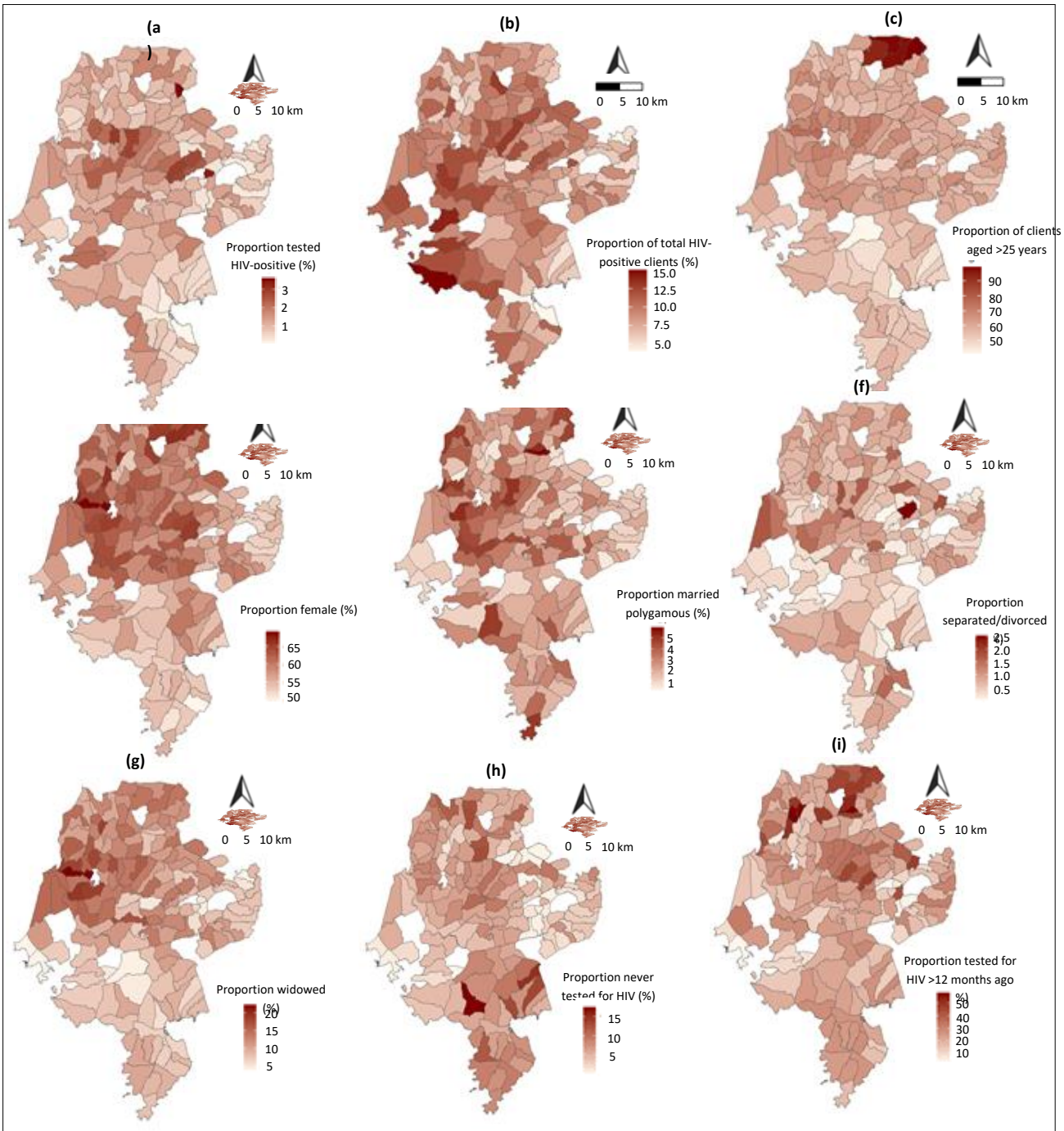
| | Enumerated, n (%) | Eligibility assessed, n (%) | Eligible for HIV testing, n (%) | Tested for HIV, n (%) | With conclusive test results, n | HIV positive, n (%) |
|---|---|---|---|---|---|---|
| Total clients | 365,798 (100%) | 355,277 (97%) | 312,223 (88%) | 268,543 (86%) | 268,153 | 2,906 (1.1%)[a] |
| Relationship to household head | | | | | | |
| Household head[b] | 136,607 (37%) | 132,622 (37%) | 111,024 (36%) | 94,506 (35%) | 94,349 | 1,432 (1.5%) |
| Spouse | 80,161 (22%) | 78,756 (22%) | 66,124 (21%) | 60,660 (23%) | 60,545 | 848 (1.4%) |
| Children ≥15 years | 110,255 (30%) | 105,999 (30%) | 99,144 (32%) | 80,781 (30%) | 80,698 | 345 (0.4%) |
| Relatives & non-relatives | 38,775 (11%) | 37,900 (11) | 35,931 (11%) | 32,596 (12%) | 32,561 | 281 (0.9%) |
| Age (median, interquartile range) | 30 (20, 47) | 30 (20, 47) | 28 (19, 46) | 28 (19, 47) | | |
| Age group (years) | | | | | | |
| 15–19 | 88,758 (24%) | 85,813 (24%) | 81,979 (26%) | 69,651 (26%) | 69,580 | 166 (0.2%) |
| 20–24 | 52,952 (14%) | 51,579 (14%) | 47,722 (15%) | 41,738 (16%) | 41,682 | 442 (1.1%) |
| 25–35 | 82,771 (23%) | 80,349 (23%) | 67,381 (22%) | 57,238 (21%) | 57,138 | 1,096 (1.9%) |
| >35 | 141,317 (39%) | 137,536 (39%) | 115,141 (37%) | 99,916 (37%) | 99,753 | 1,202 (1.2%) |
| Sex | | | | | | |
| Male | 162,628 (44%) | 156,410 (44%) | 141,011 (45%) | 114,349 (43%) | 114,187 | 1,123 (1.0%) |
| Female | 203,170 (56%) | 198,867 (56%) | 171,212 (55%) | 154,194 (57%) | 153,966 | 1,783 (1.2%) |
| Marital status[c] | | | | | | |
| Single | | | | 102,988 (38%) | 102,887 | 442 (0.4%) |
| Married monogamous | | | | 131,034 (49%) | 130,802 | 1,844 (1.4%) |
| Married polygamous | | | | 6,284 (2%) | 6,275 | 154 (2.5%) |
| Separated/divorced | | | | 1,917 (1%) | 1,913 | 100 (5.2%) |
| Widow/widower | | | | 26,317 (10%) | 26,273 | 363 (1.4%) |
| Time since last HIV test[c] | | | | | | |
| <3 months | | | | 2,521 (1%) | 2,516 | 32 (1.3%) |
| 3–12 months | | | | 183,854 (69%) | 183,606 | 1,711 (0.9%) |
| >12 months | | | | 64,870 (24%) | 64,761 | 951 (1.5%) |
| Never tested | | | | 17,298 (6%) | 17,270 | 212 (1.2%) |

[a]In addition to the new diagnoses of 2,906, a total of 27,833 previously diagnosed HIV positive clients were identified; the proportion of total HIV positive clients was 8.7% among those whose eligibility was assessed.
[b]Among household heads, 81,599 (60%) were males and 55,008 (40%) females.
[c]These variables were collected only for clients tested for HIV.
Abbreviation: n, number.




Note: Figure a) shows HIV-positive yield by age (268,153 tested with conclusive results, 2,906 HIV positive); Figure b) HIV-positive yield by sex (268,153 tested with conclusive results, 2,906 HIV positive); Figure c) HIV-positive yield by marital status (268,150 tested with conclusive results, 2,903 HIV positive); Figure d) HIV-positive yield by time since last HIV test (268,153 with conclusive results tested, 2,906 HIV positive). 95% error bar.

**Figure 4.2: The proportion of clients HIV positive (HIV positive yield) among those offered home-based HIV testing in Siaya County**




Note: Map a) shows the proportion of clients tested HIV-positive during home-based testing: number tested HIV-positive in the sub-location/ total tested in the sub-location; Map b) shows the proportion of total HIV-positive clients: number tested HIV-positive plus reporting previously HIV-positive/ total tested in the sub-location; Map c) shows the proportion of clients aged >25 years: number of clients aged >25 years/ total clients in the sub-location; Map d) shows the proportion female: number female / total clients in the sub-location; Map e) shows the proportion married polygamous: number of clients in married polygamous relationships/ total clients in the sub-location; Map f) shows the proportion separated-divorced: number of clients separated-divorced/ total clients in the sub-location; Map g) shows the proportion widowed: number of clients widowed/ total clients in the sub-location; Map h) shows the proportion never tested for HIV: number of clients never tested/ total clients in the sub-location; Map i) shows the proportion tested for HIV >12 months ago: number of clients tested >12 months ago/ total clients in the sub-location. The Geographical Information System (GIS) program, Quantum GIS version 3.6 (http://qgis.org) was used to map the different characteristics. The boundary information for sub-locations in Siaya was obtained as shapefiles from DIVA-GIS (https://www.diva-gis.org/gdata). Abbreviations: Km- kilometers.

**Figure 4.3: Maps showing sub-location distribution of different client characteristics, Siaya County**

### 4.3.3 Identification of spatial clusters of new HIV diagnoses

Sub-location level Moran's I analysis yielded an index of 0.2925 ($p<0.001$), indicating the presence of significant spatial autocorrelation of new HIV diagnoses. Nine significant sub-location clusters of higher new HIV diagnoses were identified (Figure 4.4, Table 4.13) with cluster relative risk ranging from 1.56 to 2.64, and radius ranging from 3.15 to 4.91 kilometers.

Figure 4.4 shows the location of the nine clusters of higher new HIV diagnoses. Seven of the nine clusters were located centrally in the area around, and stretching eastward and westward of Ndere town; one cluster was in the area around Ndori town, where four major roads intersect; and another was located in the south, adjacent to Lake Victoria (Figure 4.4). The sub-location cluster with the highest relative risk of 2.64 was located north-east of Ngiya town in a predominantly rural area. Significant clusters of lower new HIV diagnoses were located in the south-eastern part of the county (Figure 4.4), the area around and stretching southward of Yala town; the area south-east of Ngiya town; and the area adjacent to Lake Victoria, and stretching north, west, and south-west of Asembo town. Major roads passed through areas with clusters of higher and lower new HIV diagnoses.

Table 4.13 shows the sub-locations in each cluster, and additionally showing each cluster's radius and relative risk.

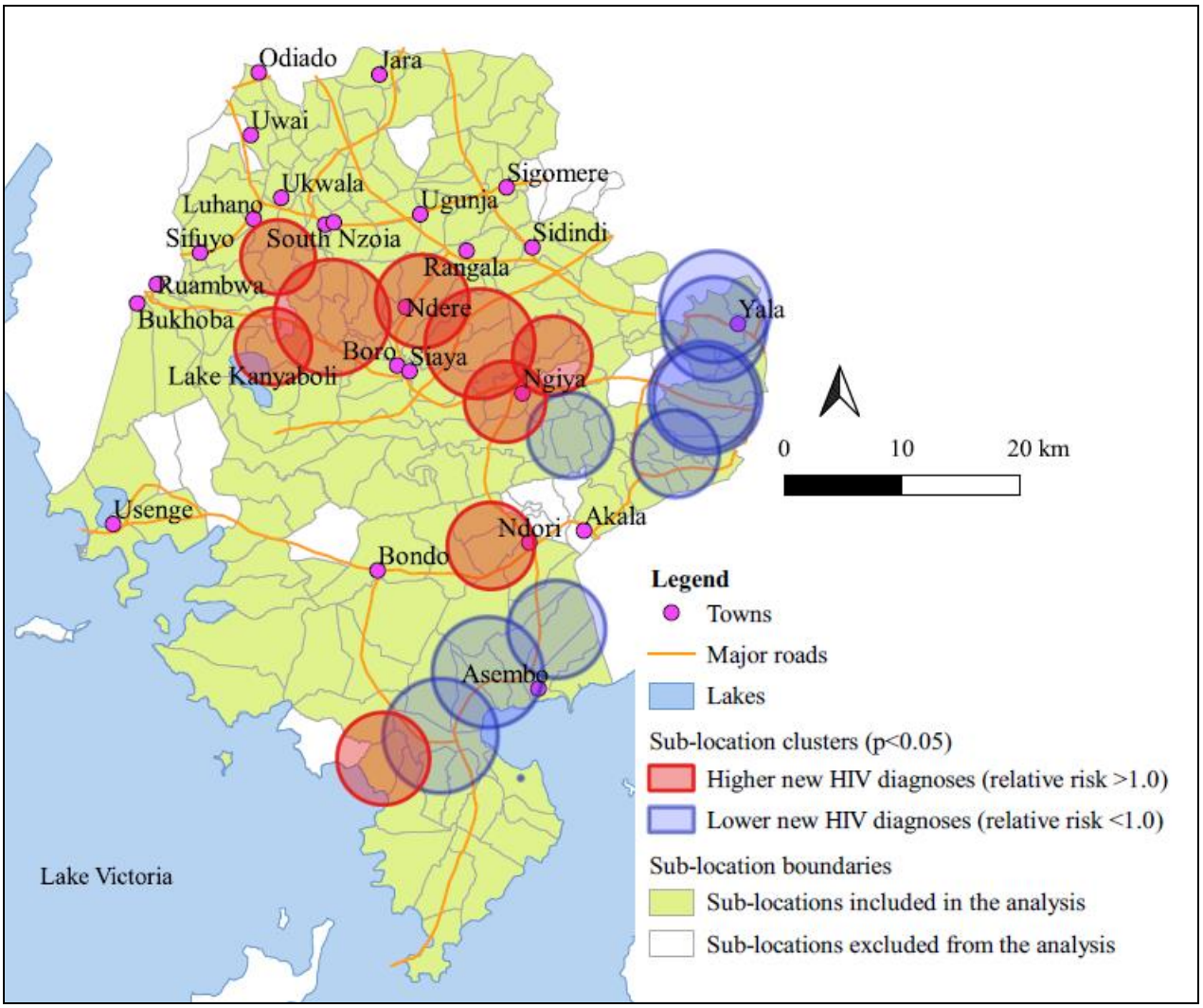



**Figure 4.4: Sub-location clusters of new HIV diagnoses from home-based HIV testing in Siaya County**

**Table 4.13: Characteristics of clusters of new HIV diagnoses[a] in Siaya County**

| Number of sub-locations in the cluster | Names of sub-locations in the cluster | Radius (kilometers) | Observed cases | Expected cases | Relative risk | Log likelihood ratio | p value |
|---|---|---|---|---|---|---|---|
| **Clusters with significant (p<0.05) higher new HIV diagnoses** | | | | | | | |
| 3 | Malunga West, Sirembe, Malunga East | 3.36 | 49 | 18.79 | 2.64 | 16.91 | **<0.001** |
| 2 | Gangu, Ojwando 'A' | 3.24 | 62 | 28.57 | 2.2 | 14.81 | **<0.001** |
| 7 | Kochieng 'A', Kodiere, Ojwado 'B', Kochieng 'B', Koyeyo, Komeny, Kalaka, Ojwando 'A' | 4.91 | 145 | 70.32 | 2.12 | 31.24 | **<0.001** |
| 5 | Komolo, Hono, Kukumu_kombewa, Nyalgunga, Koyeyo | 3.95 | 140 | 72.96 | 1.97 | 25.01 | **<0.001** |
| 4 | KomenyaKowala, KalkadaUradi, Komenya Kalaka, SimurKondiek | 3.15 | 72 | 38.93 | 1.87 | 11.4 | **0.002** |
| 7 | Ulafu, Umala, Nyalgunga, Nyamila, Olwa, Hono, Karapul | 4.65 | 197 | 111.58 | 1.82 | 27.89 | **<0.001** |
| 4 | Mur_ngiya, Olwa, Masumbi, Umala | 3.43 | 91 | 57.76 | 1.59 | 8.32 | **0.026** |
| 3 | Bar Chando, Abom, North Ramba | 3.69 | 97 | 62.91 | 1.56 | 8.12 | **0.032** |
| 2 | Kagwa, Kokwiri | 3.92 | 81 | 47.93 | 1.71 | 9.62 | **0.008** |
| **Clusters with significant (p<0.05) lower new HIV diagnoses** | | | | | | | |
| 5 | Gombe, Onyinyore, Ramula, Kambare, Uranga | 3.69 | 68 | 115.55 | 0.58 | 11.9 | **<0.001** |
| 5 | Omia Malo, OmiaDiere, Memba, South Ramba, OmiaMwalo | 4.14 | 81 | 150.33 | 0.53 | 20.11 | **<0.001** |
| 4 | Lihanda, Uranga, Marenyo, Ramula | 4.38 | 78 | 146.24 | 0.52 | 20.05 | **<0.001** |
| 6 | Bar Sauri, Nyamninia, Anyiko_yala, Jina, Nyawara, Nyandiwa_yala | 4.71 | 80 | 154.24 | 0.51 | 22.72 | **<0.001** |
| 5 | Dienya East, Nguge, Dienya West, Ulamba, Wagai West | 3.61 | 32 | 62.12 | 0.51 | 9.05 | **0.014** |
| 7 | Nyamninia, Bar Sauri, Jina, Nyandiwa_yala, Anyiko_yala, Nyawara, Marenyo | 4.41 | 99 | 192.71 | 0.5 | 29.37 | **<0.001** |
| 5 | Lihanda, Uranga, Marenyo, Ramula, Nyandiwa_yala | 4.78 | 86 | 180.17 | 0.46 | 32.17 | **<0.001** |
| 4 | Mahaya, Akom, Memba, Nyagoko | 4.68 | 56 | 119.77 | 0.46 | 21.92 | **<0.001** |
| 5 | Masala, Rachar, Akom, Kobong', Nyagoko | 4.85 | 63 | 164.62 | 0.37 | 42.97 | **<0.001** |
| 1 | Ochieng'a | 0 | 2 | 31.7 | 0.06 | 24.33 | **<0.001** |

[a]Sub-location clusters of new HIV diagnoses were mapped using SaTScan, which gradually scans a window cyclically across space, noting the number of observed and expected observations inside the window at each location, adjusting for the underlying spatial inhomogeneity of the background population. P value significant at 0.05.

### 4.3.4 Non-spatial and spatial predictors of new HIV diagnoses

In unadjusted analysis, clients aged20–24 years (uRR 4.44, 95% CI 3.73–5.33), 25–35 years (uRR 8.03, 95% CI 6.84–9.48) and >35 years (uRR 5.05, 95% CI 4.3–5.96) were more likely diagnosed HIV positive compared to those aged 15–19 years (Table 4.14). Males (uRR 0.85, 95% CI 0.79–0.92) were less likely diagnosed HIV positive compared to females. Compared to clients in monogamous marriage, clients in polygamous marriage (uRR1.74, 95% CI 1.47–2.04) or separated/divorced (uRR3.71, 95% CI 3.01–4.51) were more likely diagnosed HIV positive; while those single (uRR 0.3, 95% CI 0.27–0.34) were less likely diagnosed HIV positive. Compared to those who reported had tested for HIV 3–12 months ago, those who had never tested (uRR1.3, 95% CI 1.12–1.5) and those who had tested >12 months ago (uRR1.58, 95% CI 1.46–1.71) were more likely diagnosed HIV positive.

The non-spatial and spatial random effect multivariable models used to explore factors associated with HIV positive diagnosis are shown in Table 4.14. Of the four multivariable models explored, the convolution model that consisted of both a spatially structured and unstructured random effect model performed best with a deviation information criterion of 10,810.58. In this model, there was no association between sex (males compared to females) and HIV positive diagnosis. Clients aged 20–24 years (aRR 3.45, 95% CI 2.85–4.20), 25–35 years (aRR 4.76, 95% CI 3.92–5.81) and >35 years (aRR 2.44, 95% CI 1.99–3.00); clients in polygamous marriage (aRR 1.84, 95% CI 1.55–2.16), or separated/divorced (aRR 3.36, 95% CI 2.72–4.08); and clients never tested (aRR 2.35, 95% CI 2.02–2.72) and those who had tested >12 months ago (aRR1.53, 95% CI 1.41–1.66) were more likely to be diagnosed HIV positive.

The proportion of total HIV positive clients in a sub-location (aRR 1.3, 95% CI 1.07–1.60) was also positively associated with HIV diagnosis. Clients whose marital status was single (aRR 0.50, 95% CI 0.44–0.57) were less likely to be diagnosed HIV positive.

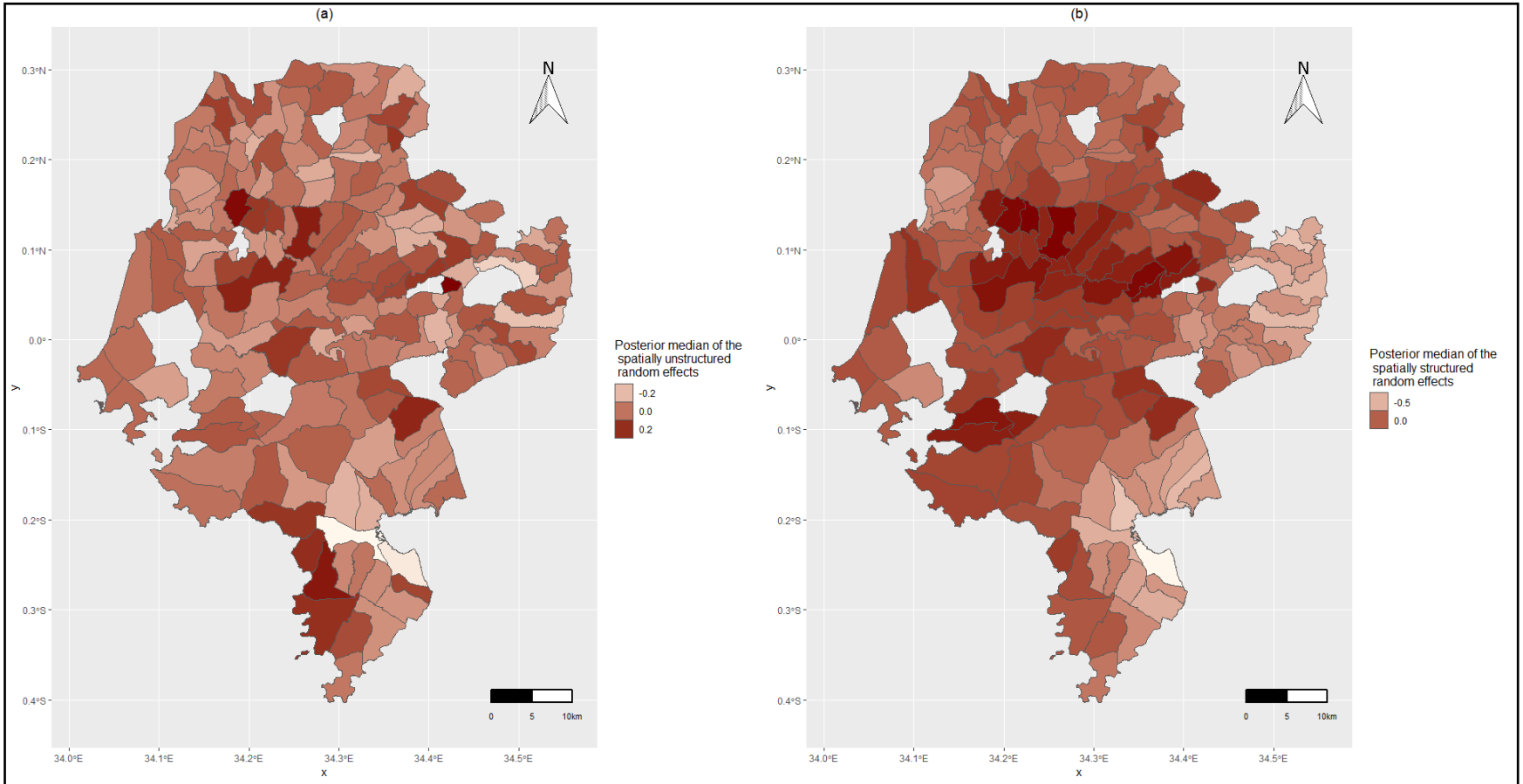
**Table 4.14: Factors associated with new HIV diagnoses in non-spatial and spatial models, Siaya County**

| Characteristic | Unadjusted relative risk (uRR) | Adjusted relative risk (aRR) | | | |
|---|---|---|---|---|---|
| | | Fixed effects only | Spatially unstructured model | Spatially structured model | Convolution spatially unstructured and structured model |
| | uRR (95% CI) | aRR (95% CI) | aRR (95% CI) | aRR (95% CI) | aRR (95% CI) |
| Age groups (years) | | | | | |
| 15–19 | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) |
| 20–24 | 4.44 (3.73–5.33) | 3.55 (2.94–4.31) | 3.46 (2.86–4.2) | 3.45 (2.85–4.19) | 3.45 (2.85–4.2) |
| 25–35 | 8.03 (6.84–9.48) | 4.78 (3.94–5.83) | 4.78 (3.93–5.82) | 4.76 (3.92–5.81) | 4.76 (3.92–5.81) |
| >35 | 5.05 (4.3–5.96) | 2.44 (1.99–3) | 2.45 (2–3.01) | 2.44 (1.99–3) | 2.44 (1.99–3) |
| Sex | | | | | |
| Female | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) |
| Male | 0.85 (0.79–0.92) | 0.95 (0.88–1.03) | 0.96 (0.89–1.04) | 0.96 (0.89–1.04) | 0.96 (0.89–1.04) |
| Marital status | | | | | |
| Married monogamous | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) |
| Married polygamous | 1.74 (1.47–2.04) | 1.86 (1.57–2.19) | 1.84 (1.55–2.17) | 1.84 (1.55–2.17) | 1.84 (1.55–2.16) |
| Separated/divorced | 3.71 (3.01–4.51) | 3.35 (2.72–4.08) | 3.37 (2.73–4.1) | 3.36 (2.72–4.08) | 3.36 (2.72–4.08) |
| Single | 0.3 (0.27–0.34) | 0.49 (0.42–0.55) | 0.5 (0.44–0.57) | 0.5 (0.44–0.57) | 0.5 (0.44–0.57) |
| Widow/widower | 0.98 (0.87–1.1) | 1.13 (0.99–1.28) | 1.1 (0.97–1.24) | 1.1 (0.97–1.24) | 1.1 (0.97–1.24) |
| Time since last HIV test | | | | | |
| <3 months | 1.36 (0.94–1.9) | 1.31 (0.9–1.81) | 1.31 (0.9–1.83) | 1.33 (0.91–1.85) | 1.33 (0.91–1.85) |
| 3–12 months | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) | 1.00 (ref) |
| >12 months | 1.58 (1.46–1.71) | 1.51 (1.39–1.63) | 1.54 (1.41–1.67) | 1.53 (1.41–1.66) | 1.53 (1.41–1.66) |
| Never tested | 1.3 (1.12–1.5) | 2.37 (2.04–2.74) | 2.35 (2.02–2.73) | 2.35 (2.01–2.72) | 2.35 (2.02–2.72) |
| Sub-location proportion of total HIV positive clients[a] | 1.61 (1.43–1.8) | 1.5 (1.34–1.68) | 1.49 (1.21–1.82) | 1.26 (1.04–1.53) | 1.3 (1.07–1.6) |
| **Random effects** | | | | | |
| Spatially unstructured precision | | | 6.01 (4.37–8.27) | | 17.39 (9.01–36.14) |
| Spatially structured precision | | | | 2.44 (1.67–3.59) | 4.97 (2.5–9.24) |
| **Model comparison** | | | | | |
| Effective number of parameters | | 13 | 125.12 | 109.09 | 113.8 |
| Deviation information criterion | | 11,153.63 | 10,816.12 | 10,811.64 | 10,810.58 |

[a]The proportion of total HIV positive clients was defined as the sum of the total new HIV positive and previously identified HIV-infected clients among those whose eligibility for HIV testing was assessed.

Abbreviations: uRR, unadjusted relative risk; aRR, adjusted relative risk; CI, credible interval.

Maps of the unstructured and structured estimated median value of the random effects for each sub location, generated from the convolution Bayesian Poisson model, are shown in Figure 4.5. The maps show the pattern of random effects, that further explain the distribution of new HIV diagnoses, over and above what is explained by the fixed effects (age group, sex, marital status, time since last HIV test and sub-location proportion of total HIV positive clients). Figure 4.5 (a) shows the pattern of posterior median unstructured random effects, not taking into account spatial autocorrelation. When spatial autocorrelation was taken into account, as shown in Figure 4.5 (b), the pattern of posterior median random effects changed, with darker areas in the central region, demonstrating higher influence of spatially correlated random effects in this area.




Note: Map a) is the estimated median value of unstructured random effects, showing residual variability of new HIV diagnoses when spatial autocorrelation was not taken into account; Map b) is the estimated median value of structured random effects, showing residual variability of new HIV diagnoses when spatial autocorrelation was accounted for. The maps were generated from the convolution Bayesian Poisson model, and mapped using ggplot2 R package (https://ggplot2.tidyverse.org). Abbreviations: Km- kilometers.

**Figure 4.5: Maps of unstructured and structured random effects of new HIV diagnosis, Siaya County**

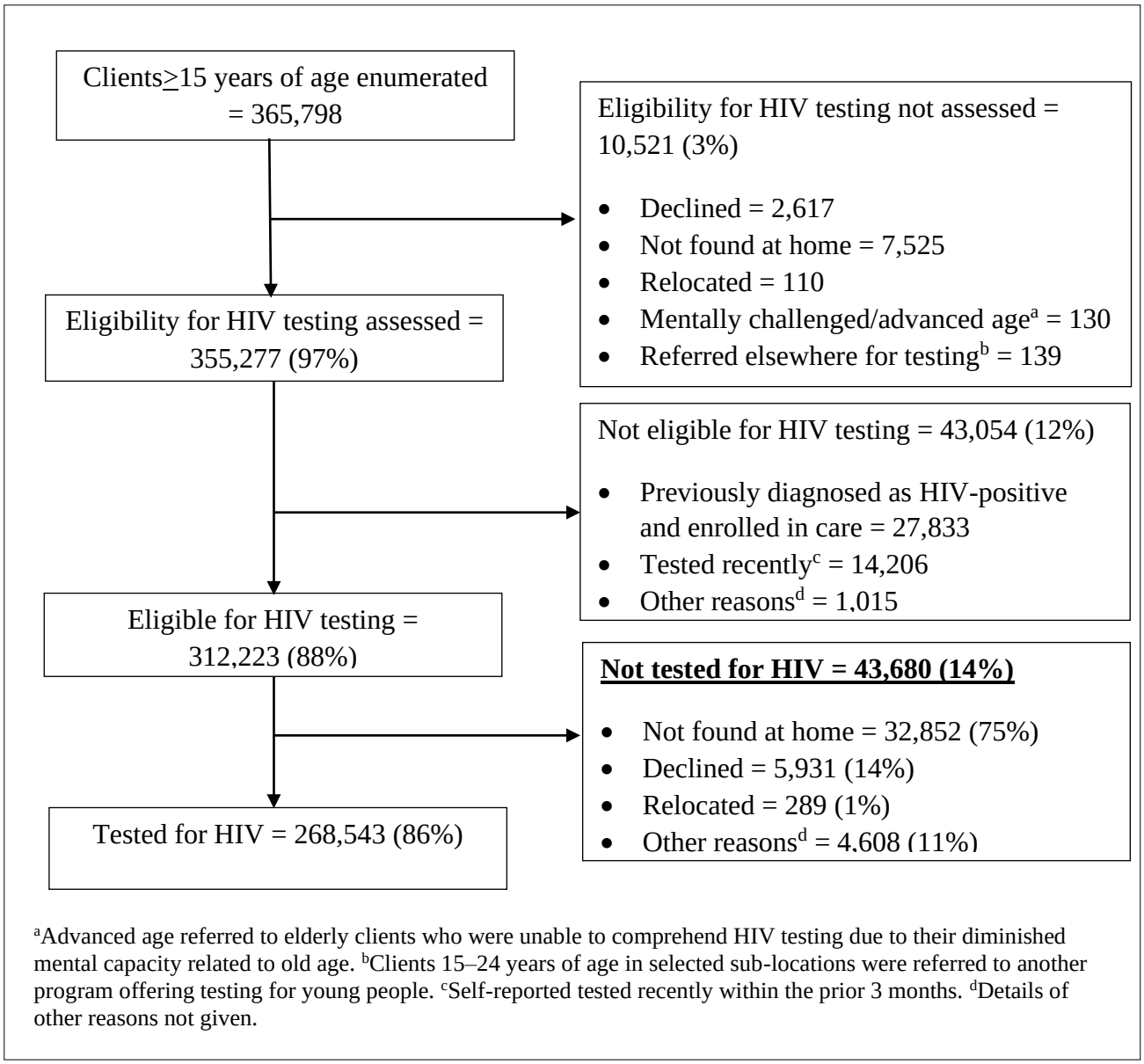
## 4.4 Results for study objective 3: The use of mapping HIV testing uptake in identifying areas with low testing uptake yet higher HIV positive yield

### 4.4.1 Characteristics of home-based HIV testing clients and overall testing uptake

A total of 365,798 clients aged ≥15 years from 136,607 households were enumerated for home-based HIV testing from the 161 Siaya administrative sub-locations included in the analysis (Figure 4.6). Overall, those enumerated had a median age of 30 years (interquartile range 20–47 years), and 203,170 (56%) were female (Table 4.15).

Of the total clients enumerated, 355,277 (97%) were assessed for HIV testing eligibility, and 312,223 (88%) were eligible for testing (Figure 4.6, Table 4.15). Among those eligible, 268,543 (86%) were tested for HIV. Of the 43,680 clients who were eligible but were not tested for HIV, the majority, 32,852 (75%) were not found at home, 5,931 (14%) declined testing, 289 (1%) had relocated their residence, and 4,608 (11%) had other reasons not given (Figure 4.6).

The 161 sub-locations had a median HIV testing uptake among eligible clients of 87% (interquartile range 82%–91%).




[a]Advanced age referred to elderly clients who were unable to comprehend HIV testing due to their diminished mental capacity related to old age. [b]Clients 15–24 years of age in selected sub-locations were referred to another program offering testing for young people. [c]Self-reported tested recently within the prior 3 months. [d]Details of other reasons not given.

**Figure 4.6: Flow diagram of clients receiving home-based HIV testing in Siaya County**

**Table 4.15: Characteristics of clients aged ≥15 years offered home-based HIV testing in Siaya County**

| Characteristic | Enumerated, n (%) | Eligibility assessed, n (%) | Eligible for HIV testing, n (%) | Tested for HIV, n (%) |
|---|---|---|---|---|
| Total clients | 365,798 (100%) | 355,277 (97%) | 312,223 (88%) | 268,543 (86%) |
| Relationship to household head | | | | |
| Household head[a] | 136,607 (37%) | 132,622 (37%) | 111,024 (36%) | 94,506 (35%) |
| Spouse | 80,161 (22%) | 78,756 (22%) | 66,124 (21%) | 60,660 (23%) |
| Children ≥15 years | 110,255 (30%) | 105,999 (30%) | 99,144 (32%) | 80,781 (30%) |
| Relatives & non-relatives | 38,775 (11%) | 37,900 (11) | 35,931 (11%) | 32,596 (12%) |
| Age (median, interquartile range) | 30 (20, 47) | 30 (20, 47) | 28 (19, 46) | 28 (19, 47) |
| Age group (years) | | | | |
| 15–19 | 88,758 (24%) | 85,813 (24%) | 81,979 (26%) | 69,651 (26%) |
| 20–24 | 52,952 (14%) | 51,579 (14%) | 47,722 (15%) | 41,738 (16%) |
| 25–35 | 82,771 (23%) | 80,349 (23%) | 67,381 (22%) | 57,238 (21%) |
| >35 | 141,317 (39%) | 137,536 (39%) | 115,141 (37%) | 99,916 (37%) |
| Sex | | | | |
| Male | 162,628 (44%) | 156,410 (44%) | 141,011 (45%) | 114,349 (43%) |
| Female | 203,170 (56%) | 198,867 (56%) | 171,212 (55%) | 154,194 (57%) |
| Marital status[b] | | | | |
| Single | | | | 102,988 (38%) |
| Married monogamous | | | | 131,034 (49%) |
| Married polygamous | | | | 6,284 (2%) |
| Separated/divorced | | | | 1,917 (1%) |
| Widow/widower | | | | 26,317 (10%) |
| Time since last HIV test[b] | | | | |
| <3 months | | | | 2,521 (1%) |
| 3–12 months | | | | 183,854 (69%) |
| >12 months | | | | 64,870 (24%) |
| Never tested | | | | 17,298 (6%) |

[a]Among household heads, 81,599 (60%) were male and 55,008 (40%) were female.
[b]These variables were collected only for clients tested for HIV.
Abbreviation: n, number.

### 4.4.2 Predictors of HIV testing uptake

In univariable analysis (Table 4.16), the following characteristics were significantly associated with higher HIV testing uptake: age20-24 (OR 1.23, 95% CI 1.19-1.27) and >35 (OR 1.16, 95% CI 1.13-1.19) years compared to age 15-19 years; and non-relatives (OR 1.48, 95% CI 0.36-1.62) compared to spouses. Characteristics significantly associated with lower HIV testing uptake included: male (OR 0.47, 95% CI 0.46-0.48) compared to female; household heads (OR 0.52,

95% CI 0.50-0.53), children (OR 0.40, 95% CI 0.38-0.41) and other relatives (OR 0.75, 95% CI 0.72-0.79) compared to spouses.

In multivariable analysis (Table 4.16), characteristics significantly associated with lower HIV testing uptake included: age 20-24 (OR 0.94, 95% CI 0.91-0.97), 25-35 (OR 0.65, CI 0.63-0.68) and >35 (OR 0.67, CI 0.65-0.70) years compared to age 15-19 years; male (OR 0.52, CI 0.51-0.53) compared to female; and household heads (OR 0.79, CI 0.76-0.82), children (OR 0.44, CI 0.42-0.45) and other relatives (OR 0.86, CI 0.81-0.90) compared to spouses. Non-relatives had a higher HIV testing uptake (OR 2.10, CI 1.92-2.30) compared to spouses. Table 4.17 shows the detailed multivariable analysis output, detailing the interactions between age and sex in association with HIV testing uptake. Overall, for all age categories, males were less likely to test compared to females. For male, younger age categories were more likely to test compared to older age categories; while for female, younger age categories were more likely to test compared to older age categories, except age 15-19 years that were less likely to test compared to age 25-35 and >35 years, and age 25-35 years that were less likely to test compared to age >35 years. For client type, children were less likely to test compared to spouses, while non-relatives were more likely to test compared to spouses.

**Table 4.16: Predictors of HIV testing uptake among clients aged ≥15 years offered home-based HIV testing in Siaya County**

| | Eligible/tested (%) | Unadjusted analysis | | Adjusted analysis | |
|---|---|---|---|---|---|
| | | OR (95% CI) | P value | OR (95% CI) | P value |
| Total clients | 312,223/268,543 (86%) | | | | |
| Age groups (years) | | | | | |
| 15-19 | 81,979/69,651 (85%) | Ref | **<0.001** | Ref | **<0.001** |
| 20-24 | 47,722/41,738 (87%) | 1.23 (1.19-1.27) | | 0.94 (0.91-0.97) | |
| 25-35 | 67,381/57,238 (85%) | 1.00 (0.97-1.03) | | 0.65 (0.63-0.68) | |
| >35 | 115,141/99,916 (87%) | 1.16 (1.13-1.19) | | 0.67 (0.65-0.70) | |
| Sex | | | | | |
| Male | 141,011/114,349 (81%) | 0.47 (0.46-0.48) | **<0.001** | 0.52 (0.51-0.53) | **<0.001** |
| Female | 171,212/154,194 (90%) | Ref | | Ref | |
| Client type in relation to household head | | | | | |
| Household head | 111,024/94,506 (85%) | 0.52 (0.50-0.53) | **<0.001** | 0.79 (0.76-0.82) | **<0.001** |
| Spouse | 66,124/60,660 (92%) | Ref | | Ref | |
| Children | 99,144/80,781 (81%) | 0.40 (0.38-0.41) | | 0.44 (0.42-0.45) | |
| Other relatives | 25,762/23,009 (89%) | 0.75 (0.72-0.79) | | 0.86 (0.81-0.90) | |
| Non-relatives | 10,169/9,587 (94%) | 1.48 (0.36-1.62) | | 2.10 (1.92-2.30) | |

P value significant at 0.05. Abbreviations: OR, odds ratio; CI, confidence interval; n, number.

**Table 4.17: Detailed multivariable analysis output of HIV testing uptake among clients aged ≥15 years offered home-based HIV testing in Siaya County**

| | OR | 95% Confidence Limits | | Comment |
|---|---|---|---|---|
| **Male vs Female by age category** | | | | |
| Male vs female at age category 20-24 years | 0.598 | 0.565 | 0.633 | Males less likely to test compared to females in all age categories |
| Male vs female at age category 25-35 years | 0.374 | 0.356 | 0.393 | |
| Male vs female at age category >35 years | 0.337 | 0.324 | 0.35 | |
| Male vs female at age category 15-19 years | 0.873 | 0.84 | 0.907 | |
| **Male: comparison of different age categories** | | | | |
| Age category 20-24 years vs 25-35 years | 1.73 | 1.65 | 1.814 | Younger age categories more likely to test compared to older age categories. |
| Age category 20-24 years vs >35 years | 1.814 | 1.728 | 1.903 | |
| Age category 20-24 years vs 15-19 years | 0.806 | 0.77 | 0.844 | |
| Age category 25-35 years vs >35 years | 1.048 | 1.011 | 1.087 | |
| Age category 25-35 years vs 15-19 years | 0.466 | 0.447 | 0.486 | |
| Age category >35 years vs 15-19 years | 0.444 | 0.425 | 0.465 | |
| **Female: comparison of different age categories** | | | | |
| Age category 20-24 years vs 25-35 years | 1.083 | 1.024 | 1.145 | Younger age categories more likely to test compared to older age categories; except age 15-19 years that are less likely to test compared to age 25-35 and >35 years; and age 25-35 years that are less likely to test compared to age >35 years. |
| Age category 20-24 years vs >35 years | 1.021 | 0.97 | 1.076 | |
| Age category 20-24 years vs >35 years | 1.176 | 1.118 | 1.238 | |
| Age category 25-35 years vs >35 years | 0.943 | 0.901 | 0.988 | |
| Age category 25-35 years vs 15-19 years | 1.087 | 1.032 | 1.144 | |
| Age category >35 years vs 15-19 years | 1.152 | 1.098 | 1.208 | |
| **Client type in relation to household head** | | | | |
| Household head vs spouse | 0.976 | 0.939 | 1.015 | Children less likely to test compared to spouses, while non-relatives more likely to test compared to spouses. |
| Children vs spouse | 0.511 | 0.49 | 0.533 | |
| Other relatives vs spouse | 1.001 | 0.949 | 1.055 | |
| Non-relatives vs spouse | 2.539 | 2.317 | 2.782 | |

Abbreviations: OR, odds ratio; CI, confidence interval.

### 4.4.3 Characteristics of clients not found at home during home-based HIV testing

Of the 43,680 clients who were eligible but were not tested for HIV, 32,852 (75%) were not found at home during home-based HIV testing. About a quarter (23%) of clients not found at home were males aged >35 years, 17% were males aged 25-35 years, 16% were males aged 15-19 years and 14% were females aged 15-19 years (Figure 4.7).

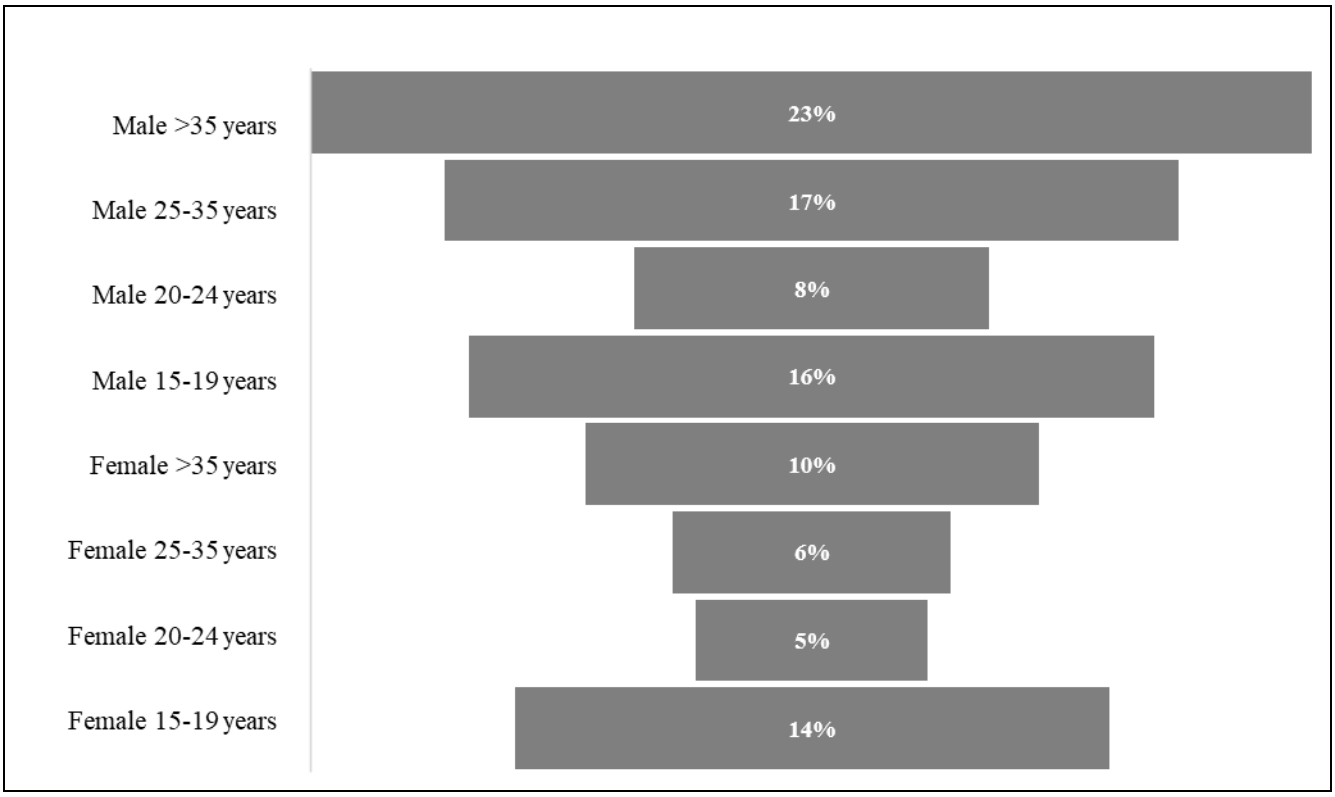



**Figure 4.7: Clients not found at home during home-based HIV testing in Siaya County, by age and sex (n=32,960)**

Compared to clients who were eligible and tested for HIV, clients who were eligible and not found at home were more likely household heads (37% vs 35%, p value <0.001), children (44% vs 30%, p value <0.001), males aged 25-35 (26% vs 22%, p value <0.001) and >35 (36% vs 33%, p value <0.001) years, and females aged 15-19 years (38% vs 23%, p value <0.001) (Table 4.18).

**Table 4.18: Characteristics of clients aged ≥15 years eligible for HIV testing but not found at home during home-based testing in Siaya County**

| | Eligible for HIV testing, but not found at home | | Eligible and tested for HIV | | P value[a] |
|---|---|---|---|---|---|
| | n | % | n | % | |
| Total | 32,852 | | 268,543 | | |
| Client type in relation to household head | | | | | |
| Household head | 12,207 | 37% | 94,506 | 35% | **<0.001** |
| Spouse | 3,585 | 11% | 60,660 | 23% | |
| Children | 14,517 | 44% | 80,781 | 30% | |
| Other relatives | 2,126 | 6% | 23,009 | 9% | |
| Non-relatives | 417 | 1% | 9,587 | 4% | |
| Male- age group (years) | | | 268,543 | | |
| 15-19 | 5,208 | 25% | 34,191 | 30% | **<0.001** |
| 20-24 | 2,690 | 13% | 17,700 | 15% | |
| 25-35 | 5,562 | 26% | 24,898 | 22% | |
| >35 | 7,574 | 36% | 37,560 | 33% | |
| Female- age group (years) | | | | | |
| 15-19 | 4,506 | 38% | 35,460 | 23% | **<0.001** |
| 20-24 | 1,764 | 15% | 24,038 | 16% | |
| 25-35 | 2,116 | 18% | 32,340 | 21% | |
| >35 | 3,432 | 29% | 62,356 | 40% | |

[a]Chi square test. P value significant at 0.05.
Abbreviation: n, number.

### 4.4.4 Characteristics of clients who declined home-based HIV testing

On the other hand, of the 43,680 clients who were eligible but were not tested for HIV, 5,931 (14%) declined testing. About a quarter (26%) of clients who declined testing were males aged >35 years, 25% were females aged >35 years; 15% were males aged 25-35 years and 12% were females aged 25-35 years (Figure 4.8).

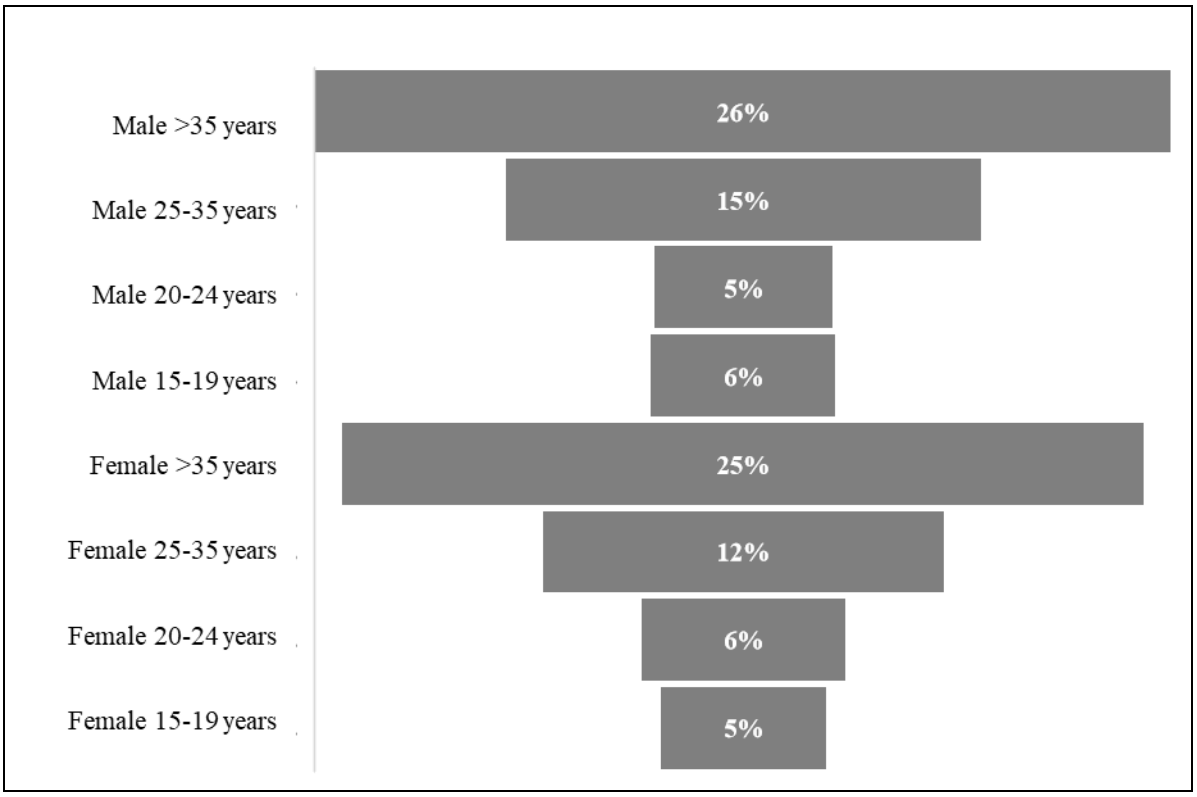



**Figure 4.8: Clients who declined testing during home-based HIV testing in Siaya County, by age and sex (n=5,931)**

Compared to clients who were eligible and tested for HIV, clients who declined testing were more likely household heads (50% vs 35%, p value <0.001), males aged 25-35 (28% vs 22%, p value <0.001) and >35 (50% vs 33%, p value <0.001) years, and females aged 25-35 (25% vs 21%, p value <0.001) and >35 (51% vs 40%, p value <0.001) years (Table 4.19).

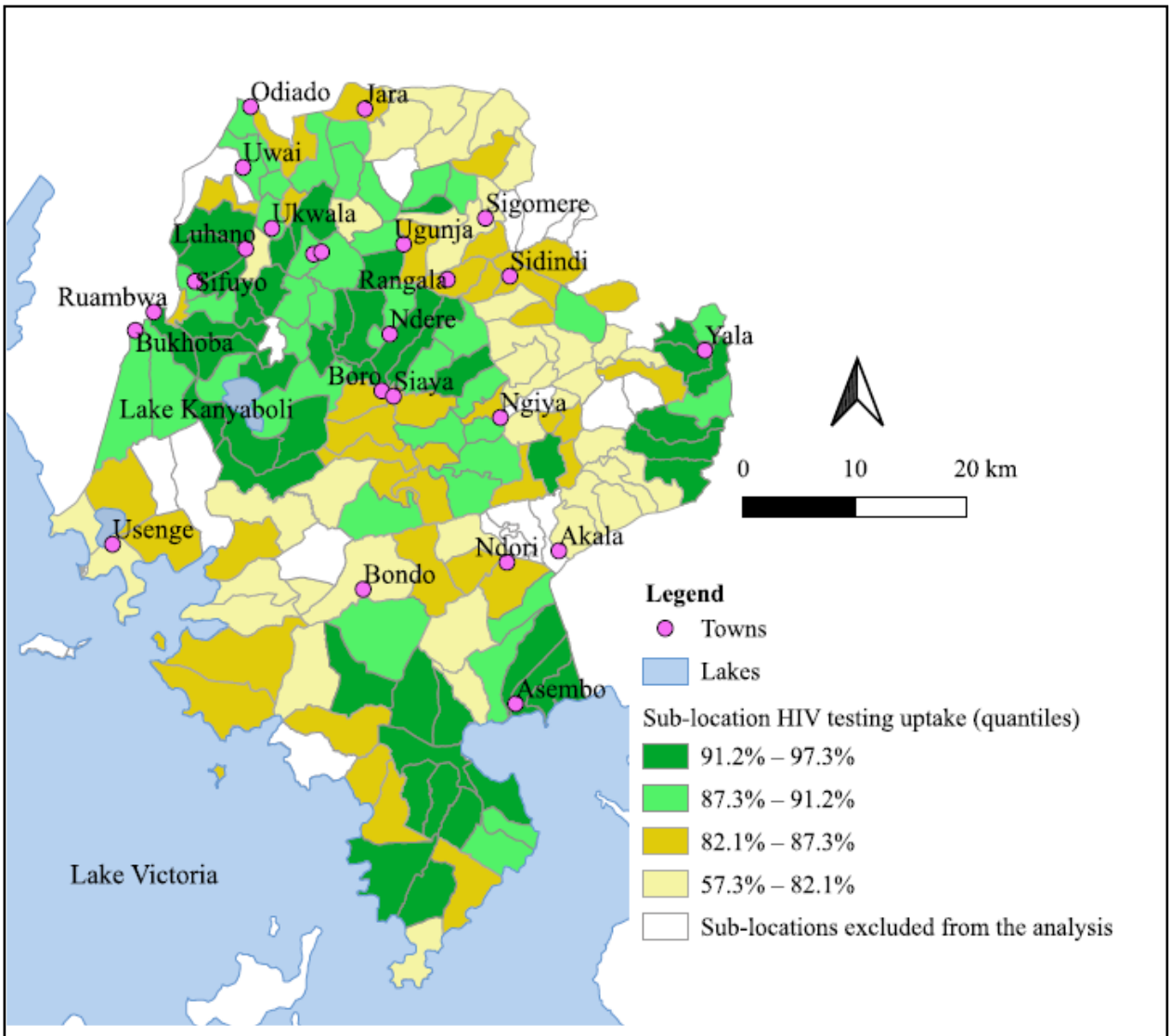
**Table 4.19: Characteristics of clients aged ≥15 years eligible for HIV testing but declined during home-based testing in Siaya County**

| | Declined testing[a] | | Accepted testing[b] | | P value[c] |
|---|---|---|---|---|---|
| | n | % | n | % | |
| Total | 5,931 | | 268,543 | | |
| Client type in relation to household head | | | | | |
| Household head | 2,949 | 50% | 94,506 | 35% | **<0.001** |
| Spouse | 1,288 | 22% | 60,660 | 23% | |
| Children | 1,283 | 22% | 80,781 | 30% | |
| Other relatives | 309 | 5% | 23,009 | 9% | |
| Non-relatives | 102 | 2% | 9,587 | 4% | |
| Male- age group (years) | | | | | |
| 15-19 | 336 | 11% | 34,191 | 30% | **<0.001** |
| 20-24 | 326 | 11% | 17,700 | 15% | |
| 25-35 | 861 | 28% | 24,898 | 22% | |
| >35 | 1,551 | 50% | 37,560 | 33% | |
| Female- age group (years) | | | | | |
| 15-19 | 302 | 11% | 35,460 | 23% | **<0.001** |
| 20-24 | 370 | 13% | 24,038 | 16% | |
| 25-35 | 728 | 25% | 32,340 | 21% | |
| >35 | 1,457 | 51% | 62,356 | 40% | |

[a]Eligible for HIV testing but declined; [b]Eligible for testing and tested for HIV; [c]Chi square test. P value significant at 0.05.
Abbreviation: n, number.

### 4.4.5 Sub-location patterns of HIV testing uptake

The map showing sub-location HIV testing uptake among eligible clients is shown in Figure 4.9. Yellow areas indicate sub-locations with the lowest HIV testing uptake range of 57-82%; while blue areas indicate sub-locations with HIV testing uptake of 82-87%.

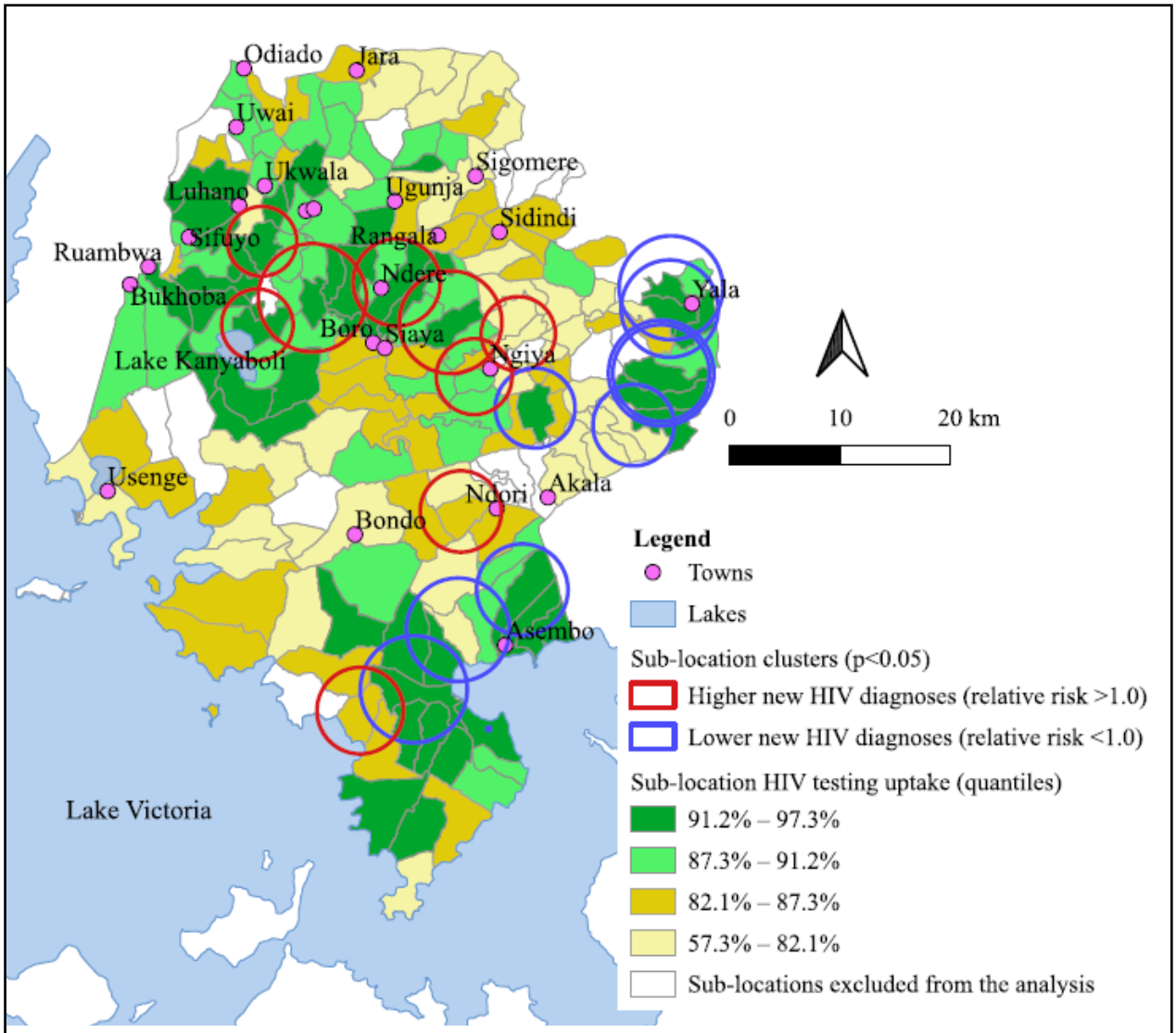



Note: a standard Geographical Information System (GIS) program, Quantum GIS version 3.6 (http://qgis.org) was used to map HIV testing uptake. The boundary information for sub-locations in Siaya was obtained as shapefiles from DIVA-GIS (https://www.diva-gis.org/gdata). For sub-location HIV testing uptake, the 0.25, 0.5, and 0.75 quantiles were used to split the data into 4 groups: category range where <25% of the observations fall; category range where 25%-50% of observations fall, category range where >50%-75% of observations fall, and category range where >75% of observations fall. Abbreviation: Km- kilometers.

**Figure 4.9: Sub-location home-based HIV testing uptake, Siaya County**

HIV testing uptake at the sub-location level overlaid with clusters of new HIV diagnoses is shown in Figure 4.10. The majority of sub-locations in clusters with higher new HIV diagnoses had high (>87%) HIV testing uptake, with exceptions observed in sub-locations located south-

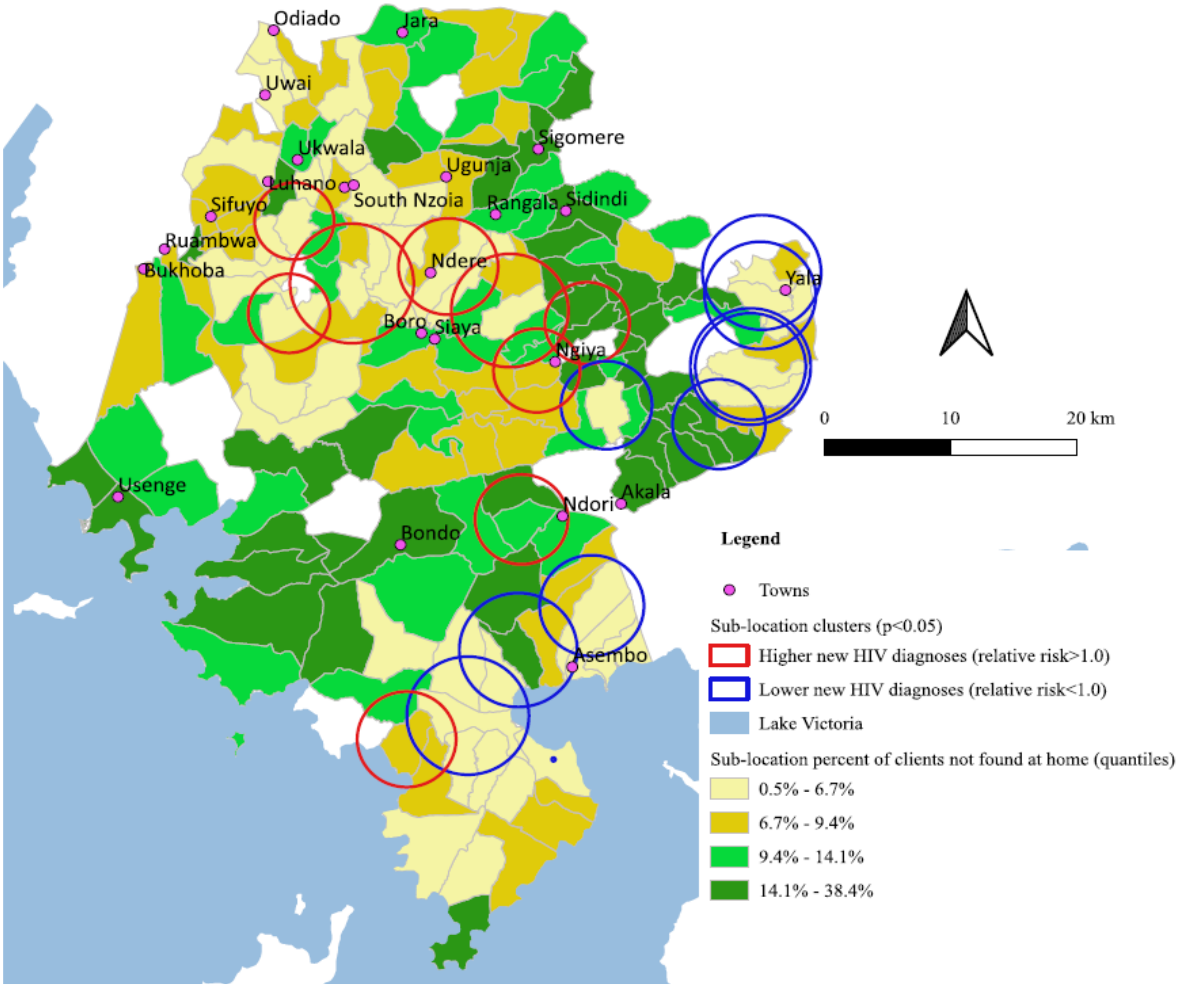
east of Luhano town; north, north-east, and east of Ngiya town; and west of Ndori town, which all had HIV testing uptake <82%.




Note: a standard Geographical Information System (GIS) program, Quantum GIS version 3.6 (http://qgis.org) was used to map HIV testing uptake, and overlay sub-location clusters of new HIV diagnoses. For sub-location HIV testing uptake, the 0.25, 0.5, and 0.75 quantiles were used to split the data into 4 groups: category range where <25% of the observations fall; category range where 25%-50% of observations fall, category range where >50%-75% of observations fall, and category range where >75% of observations fall. The clusters of new HIV diagnoses were detected using the Kulldorff's spatial scan statistic, implemented using SaTScan™ version 9.6 (http://www.satscan.org). The boundary information for sub-locations in Siaya was obtained as shapefiles from DIVA-GIS (https://www.diva-gis.org/gdata). Abbreviation: Km- kilometers.

**Figure 4.10: Sub-location home-based HIV testing uptake overlaid with clusters of new HIV diagnoses, Siaya County.**

Figure 4.11 shows the proportion of clients not found at home among those eligible for home-based HIV testing, overlaid with clusters of new HIV diagnoses. Geographic areas with clusters of higher new HIV diagnoses yet high (>9%) percent of clients not found at home were identified: north of Luhano town; south of South Nzoia town; area north-east and east of Siaya town; area west, north, and south-east of Ngiya town; area west and south of Ndori town; and area far south, near Lake Victoria.

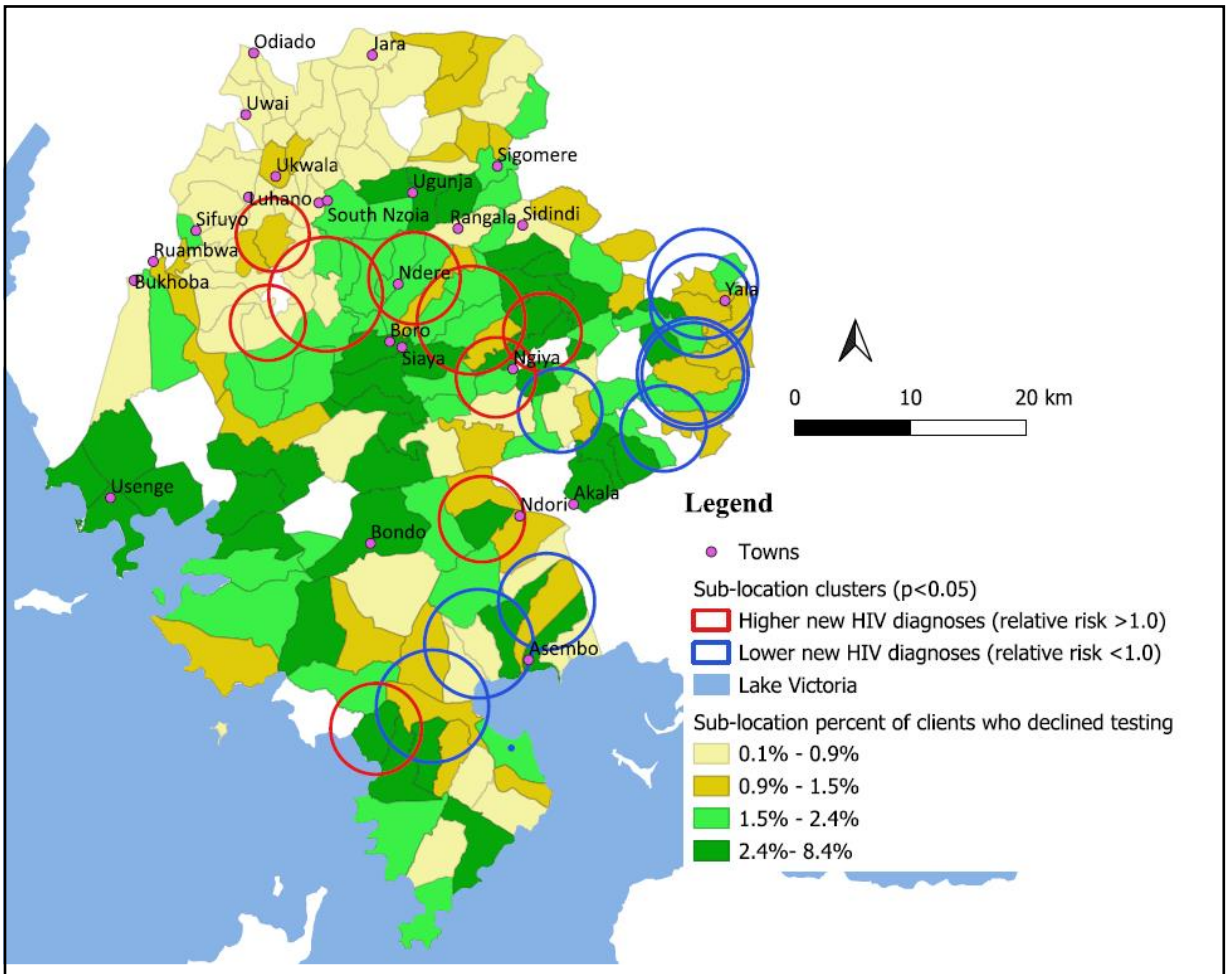



A standard Geographical Information System (GIS) program, Quantum GIS version 3.6 (http://qgis.org) was used to map the proportion of clients not found at home, and overlay sub-location clusters of new HIV diagnoses. For sub-location percent of clients not found at home, the 0.25, 0.5, and 0.75 quantiles were used to split the data into 4 groups: category range where <25% of the observations fall; category range where 25%-50% of observations fall, category range where >50%-75% of observations fall, and category range where >75% of observations fall. The clusters of new HIV diagnoses were detected using the Kulldorff's spatial scan statistic, implemented using SaTScan™ version 9.6 (http://www.satscan.org). The boundary information for sub-locations in Siaya was obtained as shapefiles from DIVA-GIS (https://www.diva-gis.org/gdata). Abbreviation- Km- kilometers.

**Figure 4.11: Sub-location proportion of clients not found at home among those eligible for home-based HIV testing, Siaya County**

Figure 4.12 shows the proportion of clients who declined testing among those eligible for home-based HIV testing, overlaid with clusters of new HIV diagnoses. Geographic areas with clusters of higher new HIV diagnoses yet high (>1.5%) percent of clients who declined testing were

identified: area stretching from west of Ndere town to area around Ngiya town; area west of Ndori town; and area far south, near Lake Victoria.




A standard Geographical Information System (GIS) program, Quantum GIS version 3.6 (http://qgis.org) was used the proportion of clients who declined testing, and overlay sub-location clusters of new HIV diagnoses. For sub-location percent of clients who declined testing, the 0.25, 0.5, and 0.75 quantiles were used to split the data into 4 groups: category range where <25% of the observations fall; category range where 25%-50% of observations fall, category range where >50%-75% of observations fall, and category range where >75% of observations fall. The clusters of new HIV diagnoses were detected using the Kulldorff's spatial scan statistic, implemented using SaTScan™ version 9.6 (http://www.satscan.org). The boundary information for sub-locations in Siaya was obtained as shapefiles from DIVA-GIS (https://www.diva-gis.org/gdata). Abbreviation- Km- kilometers.

**Figure 4.12: Sub-location proportion of clients who declined testing among those eligible for home-based HIV testing, Siaya County.**

# CHAPTER FIVE
# DISCUSSION

## 5.1 Introduction to discussion

This study found that the following three strategies evaluated: a HIV predictive risk-score screening algorithm, geospatial analysis of new HIV diagnoses, and mapping HIV testing uptake, successfully identified sub-populations and granular geographic areas with higher HIV positive yield, useful to inform the targeting of HIV testing for maximal epidemiologic impact and efficient use of resources. The findings for the three study objectives are discussed below.

## 5.2 Discussion for study objective 1: The use of a HIV predictive risk-score screening algorithm

This study developed and validated a HIV predictive risk-score algorithm, derived from a set of socio-demographic and behavioral characteristics. The algorithm successfully identified outpatient sub-populations that have higher risk of HIV infection, to whom HIV testing should be targeted for efficiency. The final predictive screening algorithm assigned scores to each of the following characteristics: age categories 35–39 years or 40–44 years; manual/domestic or trade/sales/service occupation; polygamous marriage, separated/divorced or widowed; in the prior 12 months having ≥2 sexual partners; reporting treatment for an STI; and never been tested for HIV or having a negative HIV test result >12 months prior. This study found that, in the overall algorithm, targeting HIV testing to patients meeting risk-scores of ≥10, would dramatically reduce (by about 75%) the number of patients tested, and still identify about 50% of those HIV infected. The predictive algorithm accounted for a high proportion of the variability of HIV prevalence in the development ($R^2$ 0.89) and validation ($R^2$ 0.88) study populations. The algorithm's ability to discriminate between individuals with, and without, HIV infection in the outpatient setting was modest (AUC of 0.69 for both the development and validation datasets).

Few studies have evaluated the use of HIV predictive algorithms in outpatient settings. Three studies in the United States and Spain (Elías et al., 2016; Haukoos et al., 2012; Lyons et al., 2013) evaluated HIV predictive algorithms among outpatient attendees. The United States and Spain, where the studies were conducted, have concentrated epidemics, and HIV risk factors are mostly related to sexual relations within specific affected sub-populations. The HIV risk factors differ from those in the sub-Sahara Africa region, that largely has a generalized HIV epidemic, and risk factors are mostly related to heterosexual relationships in the general population. By the time this study was conducted (2021), no HIV screening algorithms or studies were found that had evaluated algorithms among adults in the outpatient setting in sub-Sahara Africa. This study therefore presents results from the development and evaluation screening algorithm that is context specific to the sub-Saharan Africa region.

This study's HIV predictive risk-score algorithm consisted of simple variables, which were collected within a routine health care delivery setting, showing the feasibility of implementation. This study's AUC was 0.69, demonstrating a modest performance in discriminating individuals with HIV infection. The discrimination performance of the algorithm was compared with that of other studies conducted among outpatient attendees. The "Denver HIV Risk Score", was developed using patient data from a metropolitan sexually transmitted disease clinic in Denver, Colorado, and has been validated in several settings including in the outpatient (Haukoos et al., 2013; Haukoos et al., 2015; Hsieh, Haukoos, & Rothman, 2014). The Denver risk-score algorithm score included age, gender, race/ethnicity, sex with a male, vaginal intercourse, receptive anal intercourse, injection drug use, and past HIV testing. The Denver algorithm reported a higher AUC (range of 0.75–0.85), likely because the algorithm was derived using data from STI clinic attendees who likely have higher HIV risk compared to the outpatient population

in this study. The lower AUC noted in this study likely reflects more widespread distribution of HIV-risk factors among persons accessing health facility out-patient services even in the setting of a generalized HIV epidemic. Another study done in the United States (Lyons et al., 2013), compared targeted screening and universal testing, among patients attending the emergency outpatient department. The study concluded that targeted screening, even when fully implemented with maximally permissive selection, offered no important increase in positivity rate, or decrease in tests performed; universal screening diagnosed more cases, because more were tested, despite a modestly lower consent rate. A study done in Spain (Elías et al., 2016) developed a Spanish-structured HIV risk of exposure and indicator conditions (RE&IC) questionnaire. The HIV RE&IC questionnaire had a high sensitivity of 100% to predict HIV infection, but a much lower specificity of 49%. Another study published recently in 2022, after this study's publication, was conducted in Malawi (Moucheraud et al., 2022), and used exit survey data collected at outpatient departments to develop a screening algorithm. The study found that suspect STIs and having ≥ 3 sexual partners were associated with HIV positivity, but had weak sensitivity and specificity. The full tool (using the optimal cut-off score of ≥ 3) achieved 55.6% sensitivity and 84.9% specificity for HIV positivity; the reduced tool (optimal cut-off score ≥ 2) achieved 59.3% sensitivity and 68.5% specificity; and standard of care had 77.8% sensitivity and 47.8% specificity. The study concluded that the screening tool for HIV testing in the outpatient department did not offer clear advantages over the standard of care. These studies demonstrate variation in the performance of HIV predictive algorithms evaluated in outpatient settings in different countries/communities, even within the sub-Saharan Africa region. This emphasizes the need to have predictive algorithms that are designed for specific contexts.

HIV predictive algorithms assessed for use in other populations (children, adolescents, key populations, and women) have also demonstrated variation in their performance. Among children 0-15 years, prospective and cross-sectional studies done in settings of generalized HIV epidemics (Clemens et al., 2020), reported a sensitivity of screening tools ranging from 71% to 96%, and specificity ranging from 25% to 99%. Evaluation of risk score algorithms among adolescents in sub-Sahara Africa found modest to low sensitivity (range of 56% to 74%), modest specificity (range of 75% to 80%), and modest to low AUC (range of 0.55 to 0.65) (Bandason et al., 2018; Ferrand et al., 2011). Despite the varied performance of algorithms among children and adolescents, studies have reported good performance of screening tools/algorithms among high-risk populations. A large study in the United States that included Black MSM showed good discrimination, with C-statistics ranging from 71.0 to 73.1. A study done in Beijing China, calibrated well; with bootstrap-corrected c-indices ranging between 0.70 and 0.71. A multi-country study done in Uganda, Tanzania, and Mozambique among FSWs, MSM and fisher folk, showed good predictive ability with an AUC of 0.70. These results further demonstrate that predictive algorithms have better performance among populations with overall higher HIV-risk.

Although many of the studies described above, have shown variation in the performance of HIV predictive screening algorithms evaluated for use in different populations and different regions including in sub-Sahara Africa, this study's predictive algorithm, generated and assessed in the outpatient setting in Kenya (a sub-Sahara country), was found useful in identifying sub-populations with higher HIV risk. When applied, the algorithm would reduce the number of HIV tests done by about 75%, and still identify about 50% of those HIV positive; and showed a modest discrimination performance (AUC 0.69).

The performance of this study's algorithm was further assessed separately among female and male. Among female, the proportion of variability in HIV prevalence accounted for by the final model/algorithm was high ($R^2$ of 0.87 in the development and 0.95 in the validation datasets), and varied among male ($R^2$ of 0.69 and 0.85 in the development and validation datasets, respectively). The performance of the algorithm in discriminating patients with, and without, HIV infection was modest among female (AUC of 0.66 for both the development and validation datasets). This modest AUC was comparable to what was reported in a study that used data from 3 randomized trials among women in Africa, that had AUCs ranging between 0.58 and 0.71 (Balkus et al., 2016). The AUC among males in this study, was however somewhat higher (AUC of 0.76 and 0.71 for the development and validation datasets, respectively). During this study's literature review, no studies were found that have assessed the use of algorithms among men, except those among MSM. The higher AUC seen among males in this study, may be reflective of the strong contribution of circumcision status to the algorithm, as this variable was assessed only among males. Although this study highlights variation in the performance of gender-specific algorithms, the majority of the HIV-risk factors included in the final models were similar for both sexes. The use of a single overall algorithm may, therefore, be appropriate and likely more feasible to implement in the field.

This study found that targeted HIV testing using the three highest risk-score categories (risk scores of $\geq$10) in the overall algorithm, would dramatically reduce (by about 75%) the number of patients tested; however, this approach would miss the diagnosis of approximately 50% of HIV infected individuals accessing health facilities, making the use of the algorithm inferior to universal testing. Even for the gender-specific algorithm among males, which had superior discrimination performance as compared to the overall algorithm, targeted HIV testing using the

three highest risk-score categories (risk scores of ≥13 for males) would reduce the number of patients tested by one half, and miss the diagnosis of approximately 14% of HIV infected individuals. The algorithm's use should, therefore, be considered in settings where resource or other logistical constraints necessitate targeted testing, and should be coupled with other HIV testing strategies recommended by the WHO (World Health Organization, 2015, 2016).

The predictors included in the risk-score algorithm are consistent with those shown in other studies to be associated with higher risk of HIV infection. The pattern of HIV prevalence by age and sex in this study is consistent with national surveys in Kenya (National AIDS STI Control Programme Ministry of Health Kenya, 2013). Similar to this study, other studies have shown that the following factors are associated with higher risk of HIV infection: polygamous marriage (Bove & Valeggia, 2009), widowed status (Amornkul et al., 2009; Kimani et al., 2013; Oluoch et al., 2011; Tenkorang, 2014), or separated/divorced status (Kimanga, Ogola, & Umuro, 2014; Kimani et al., 2013; Oluoch et al., 2011; Tenkorang, 2014); having multiple sexual partners (Amornkul et al., 2009; Kimanga, Ogola, & Umuro, 2014; Oluoch et al., 2011; Pettifor et al., 2005); having a new sexual partner (Amornkul et al., 2009; Pettifor et al., 2005); having an STI (Amornkul et al., 2009; Pettifor et al., 2005); and uncircumcised status among men (Auvert et al., 2005; Bailey et al., 2007; Kimanga, Ogola, & Umuro, 2014; Oluoch et al., 2011; Wawer et al., 2009). Some studies have shown an association between HIV risk and higher socioeconomic status/employment/having income (Amornkul et al., 2009; Kapiga, Lyamuya, Vuylsteke, Spiegelman, Larsen, & Hunter, 2000; Msisha, Kapiga, Earls, & Subramanian, 2008), others have shown an association with low socioeconomic status (Farmer, 2001), while others have demonstrated a mixed association (Hargreaves et al., 2002; Wojcicki, 2005) or no association (Ayisi, Van-Eijk, Kuile, Kolczak, Otieno, & Misore, 2000). Although in this study

socioeconomic status was not assessed directly, the study found manual/domestic, or trade/sales/service occupations, were associated with higher risk of HIV infection. This might be explained by an interplay between source of income and behavior, including increased opportunity for social interaction and travel. The study results were consistent with program data from western Kenya, which found that patients who had never been tested for HIV, or had a negative HIV test result >12 months prior were more likely HIV positive (Joseph et al., 2019). This is because this study similarly used routine program data. Most patients (95%) had been tested for HIV within the previous 12 months, reflecting intensified HIV testing efforts to increase ART coverage in the study region (United States President's Emergency Plan for AIDS Relief, 2015, 2016, 2017; US President's Emergency Plan for AIDS Relief, 2017). Unlike studies that have shown that alcohol use (Kalichman, Simbayi, Jooste, Vermaak, & Cain, 2008; Zablotska et al., 2006), intimate partner violence (Annie, 2011), and money in exchange for favors or being a key population (Choudhry, Ambresin, Nyakato, & Agardh, 2015; Grabowski, Burke, Nakiigozi, Nalugoda, Ssekubugu, & Chang, 2018; Kenya National AIDS Control Council, 2009) are associated with higher risk of HIV infection, these factors were not significant in this study. This is possibly owing to these variables being under-reported or being less prevalent in the study population of general outpatient attendees. Although other studies have demonstrated an association of race/ethnicity with HIV infection (Dube et al., 2018), likely due to cultural, socioeconomic, and other disparities that exist between people of different races and ethnic groups, mostly seen in developed countries, this association has not been shown by studies conducted in Kenya where these disparities may not be distinct between different ethnic groups. This factor was not evaluated in this study. Some studies have demonstrated an association between HIV prevalence and higher education level (Hargreaves & Glynn, 2002;

National AIDS STI Control Programme Ministry of Health Kenya, 2013), while others an association with lower education level (Hargreaves & Glynn, 2002; National AIDS and STI Control Programme, 2008). Additionally, inconsistent condom use (A. Pettifor, Rees, HV., Kleinschmidt, I., *et al*., 2005) and having an HIV infected sexual partner (Eshleman, 2017) have been associated with higher HIV infection. Education level, condom use and having a HIV infected sexual partner were not included as potential HIV predictor variables in this study, hence their effect was not assessed, and should be considered in future studies. This study, however, included the majority of behavioral characteristics that have been demonstrated to be associated with higher HIV infection in the study setting.

Behavioral risk data were collected by trained counselors at a private space, to facilitate patient privacy and reduce social desirability bias. However, a comparison of this study's patient characteristics with results from the most recent (2014) Kenya Demographic and Health Survey suggests patients might have under-reported certain variables. The survey reported that 1.7% of females and 22% of males in the study region use alcohol (Kenya National Bureau of Statistics, 2014), suggesting that the proportion of patients in this study who reported having sex under the influence of alcohol (2%) is likely an underestimate. Similarly, whereas the survey results showed that nationally 7.8% of women and 2.3% of men experience sexual violence (Kenya National Bureau of Statistics, 2014), this study found that 2% of patients reported being coerced to have sex in the prior 12 months, also likely an underestimate. The proportion of patients who reported having sex in exchange for money/favors (3%) in this study is, however, comparable to the national survey findings (Kenya National Bureau of Statistics, 2014).

This study had some limitations. Although the development of the algorithm derives strength from using data from five health facilities located across three counties, data used for external

validation was from a facility located in the same region. The algorithm should therefore be externally validated in other regions and settings, and the impact of its use evaluated.

### 5.3 Discussion for study objective 2: The use of geospatial analysis of new HIV diagnoses in identifying areas with higher HIV positive yield

This study successfully identified granular-geographic clusters of new HIV diagnoses using geospatial analysis. Although the HIV epidemic in Siaya is generalized, this study identified nine sub-location clusters in which the number of new HIV diagnoses observed was 1.56 to 2.64 times higher than expected.

This study aimed to conduct geospatial analysis of new HIV diagnoses to the smallest possible geographic unit. Although analysis at the village-unit was desired, it was not statistically feasible, as the total population of the village, the number of clients tested for HIV, and those identified as HIV positive were too small for meaningful statistical analysis. Analysis at sub-location level was therefore conducted, which resulted in mapping of clusters of new HIV diagnoses to 5-kilometer radius areas, useful to inform targeting of HIV testing to granular geographic units for efficiency. Other studies, including in Kenya, have used geospatial analysis to map HIV-related factors to larger geographic units. A study using routine facility-level HIV testing data in Kenya identified facility clusters, at a radius of $\leq$50 kilometers, of newly diagnosed HIV positive persons across counties with differing HIV burden (Waruru et al., 2021b). Another study conducted in Kenya, mapped HIV prevalence clusters to 100-kilometer radius areas (Waruru et al., 2018); and a study that used data from 20 sub-Saharan Africa countries, mapped spatial distribution of HIV infection to 100-kilometer radius areas (Cuadros, Awad, & Abu-Raddad, 2013a). A study done in KwaZulu-Natal in South Africa, uniquely mapped HIV prevalence and seroconversion to describe the contribution of high-risk locations in

the overall HIV transmission network, using 3-kilometer Gaussian kernels (Cuadros et al., 2022). This study, therefore, uniquely mapped clusters of higher new HIV diagnoses, to 5-kilometer radius areas, using data collected routinely from home-based testing, to inform granular targeting of HIV testing.

Geospatial analysis has been further used to describe spatial patterns in HIV prevalence, incidence, HIV treatment cascade, and access to health-care services. Studies that have described patterns of HIV prevalence include: the description of case distribution trends and population characteristics at census tract level (Hixson et al., 2011),the description of geographic variation of HIV prevalence across sub-Sahara African countries (Cuadros, Awad, & Abu-Raddad, 2013a), identification of geographic clusters of HIV prevalence (Waruru et al., 2018), description of microgeographic patterns and clustering of HIV infections in a high prevalence setting (Tanser et al., 2017), and mapping and characterization of high prevalence areas of young adults (Bulstra et al., 2020). Studies that have used geospatial analysis to describe spatial patterns of HIV incidence include: identification of hot-spot areas in the growth of new HIV infections (Zhu et al., 2021), and generation of surface maps of HIV prevalence and HIV seroconversion (Cuadros et al., 2022). One study spatially characterized the HIV treatment cascade (Eberhart et al., 2013), and another generated high-resolution maps of underserved areas where people cannot access the closest health care facilities within appropriate travel time in sub-Sahara Africa (H. Kim et al., 2021).

In Siaya, clusters of higher new HIV diagnoses were found in areas around specific towns, around major roads, near a major road intersection and adjacent to a beach. Other studies have described the clustering of higher HIV prevalence and incidence around similar ecological factors (lake/river, major road/highway, economic hub, or in highly productive agricultural

zones) (Waruru et al., 2018). The effect of mobility and migration intensity in predicting HIV acquisition risk in high-incidence communities near a major road has been described in South Africa (Tanser et al., 2017). The clustering around ecological features observed in this study suggests that population-level factors related to the ecological features, including socioeconomic, mobility and geographic factors, may influence the clustering of new HIV diagnoses. Surprisingly, however, the sub-location cluster with the highest relative risk was in a predominantly rural area with no prominent ecological features. Furthermore, several sub-locations around towns and major roads had clusters of lower new HIV diagnoses, suggesting that other unidentified factors, additionally influence the distribution of new HIV diagnoses.

This study uniquely used a Bayesian model to enable the assessment of individual and spatial-level associations of new HIV diagnoses in a spatially integrated framework. Spatial effects influenced the distribution of new HIV diagnoses, influencing the degree of association of individual-level factors, and further influencing the pattern of random effects (the distribution of new HIV diagnoses not explained by factors in the Bayesian model). The spatial Bayesian model found that clients in polygamous marriage and those separated/divorced were more likely diagnosed HIV positive, likely due to their higher risk of HIV infection as shown in other studies (Adeokun & Nalwadda, 1997; Bove & Valeggia, 2009; Kimanga, Ogola, & Umuro, 2014; Oluoch et al., 2011; Tenkorang, 2014). Separated/divorced women have been shown to have a higher risk of HIV (Boileau et al., 2009), as these women may seek new sexual relationships that put them at higher risk of HIV, or HIV infection may have contributed to the divorce/separation (Porter et al., 2004). Although several studies have documented a correlation between widowhood and higher HIV infection (Amornkul et al., 2009; Oluoch et al., 2011; Tenkorang, 2014), a significant association between widowed individuals and HIV positive diagnosis was

not observed in this study. Similar to findings observed in facility-based testing (Joseph et al., 2019), individuals who had never tested for HIV, and those tested >12 months prior, were more likely to be diagnosed HIV positive. The association between increasing age and higher likelihood of HIV diagnosis found in this study is consistent with higher HIV prevalence observed in older age groups (National AIDS STI Control Programme Ministry of Health Kenya, 2013). Although other studies have shown that men have lower HIV prevalence compared to women, this study's spatial model did not find a significant association between HIV positive diagnosis and sex. The association observed between higher proportion of total HIV positive clients in a sub-location and higher new HIV diagnoses suggests these areas likely have a relatively high number of undiagnosed people living with HIV and ongoing local HIV transmission. Random effects or additional factors beyond those included in the Bayesian model, influenced the distribution of new HIV diagnoses. This points to the importance of other factors, likely other individual or population-level factors (including geographic, economic, or social), that influenced the pattern of new HIV diagnoses.

Home-based HIV testing conducted in Siaya between May 2016 and July 2017 achieved high (86%) HIV testing uptake among eligible individuals; and was comparable to the testing uptake (64% to 99%) reported in other home-based testing programs in sub-Saharan Africa (Sabapathy et al., 2012). The proportion of new HIV diagnoses was low (1.1% HIV positive yield), slightly lower than that observed in outpatient HIV testing services (1.3% yield) in this setting (Joseph et al., 2019). The low yield observed is likely due to a diminishing number of undiagnosed people living with HIV in the general population, and further highlights the importance of granular spatial analysis to better target HIV testing programs.

The number of individuals aged ≥15 years enumerated for home-based testing in the 161 sub-locations included in the analysis (365,798 clients) were compared with the 2016/2017 corresponding projected population (435,727 individuals). The projected population was derived using 2009 (Kenya National Bureau of Statistics, 2009) and 2019 (Kenya National Bureau of Statistics, 2019) Kenya population census reports. From this, it was estimated that the majority (~84%) of residents aged ≥15 years in the 161 sub-locations included in the study analysis were enumerated for home-based testing.

This study had some limitations. First, the results did not represent the whole of Siaya County, as data for 18 sub-locations were excluded; the study did, however, include the majority (90%) of sub-locations in the county. Second, several limitations were encountered owing to the use of data collected during the routine provision of home-based HIV testing services, namely: HIV testing procedures were those set for the routine home-based testing program; during enumeration, household residents who reported they would be away for more than one-month following enumeration were excluded, which might have reduced representation of adolescents in boarding schools/colleges; data were not available to verify the number of households in each sub-location enumerated; and variables included in the analysis of factors associated with new HIV diagnoses were limited to those collected during the routine provision of home-based HIV testing services, and therefore it was not possible to explore other variables likely associated with new HIV diagnoses. Third, per Kenya Ministry of Health guidelines, the assessment of HIV testing eligibility relied on self-reported previous HIV testing, which can be unreliable (A. Kim et al., 2016).

Finally, despite literature showing the usefulness of geospatial analysis in informing geographic-targeting of HIV interventions (Anderson et al., 2014; Lilian, Grobbelaar, Hurter, McIntyre,

Struthers, & Peters, 2017; Waruru et al., 2021b), geospatial analysis is not routinely used in public health programs. This study demonstrates the feasibility of using routine HIV testing data for geospatial analysis, to identify granular (≤5 kilometers) geographic areas to target HIV testing and other interventions. Although this study used data collected during the provision of routine home-based testing, other routinely available HIV testing data (e.g., provider-initiated testing and counseling data at health facilities, data from partner HIV testing services/index testing, antenatal clinic data, etc.) could be used in a similar manner. It is recommended that countries and programs should integrate geospatial analysis into routine public health program data analysis and use, to inform targeting of interventions to more granular geographic units for maximal epidemiologic impact and efficient resource allocation.

### 5.4 Discussion for study objective 3: The use of mapping HIV testing uptake in identification of areas with low testing uptake yet higher HIV positive yield

By mapping HIV testing uptake, this study successfully identified granular areas of low testing uptake yet higher HIV positive yield, useful to target HIV testing for efficiency. Additionally, by mapping the sub-location proportion of clients who were not found at home during home-based testing, and those who declined testing, areas with higher HIV positive yield yet with a high proportion of clients not found at home or who declined testing were identified, again useful to inform targeting of HIV testing services.

Similar to other studies, this study found variation in home-based testing uptake by granular geographic units/ sub-locations. Areas of low testing uptake were identified, with quantile testing uptake as low as 57% to 82%. By overlaying the maps with areas of higher new HIV diagnoses, areas of low testing uptake yet have higher new HIV diagnoses were identified. Targeting these areas with HIV testing services would be critical, to increase testing uptake in an efficient

manner. Several studies have similarly mapped HIV testing uptake. A study in Ghana developed an HIV testing prevalence surface map using spatial interpolation techniques to identify geographical areas with low and high HIV testing; the surface map further revealed intra-regional level differences in HIV testing estimates (Nutor et al., 2021). A study done in Ethiopia (Alem, Liyew, & Guadie, 2021) described factors and geographic variation of home-based testing uptake. The study found that the spatial patterns of home-based testing uptake were non-random (Global Moran's I = 0.074, *p* value< 0.001). Forty-seven primary clusters were identified that were located in the entire Somali region with a relative likelihood of 1.50. Few studies have described factors that may influence a community's or geographic region's HIV testing uptake. A study done in Ethiopia noted that exposure to education materials, HIV-related stigma, and high percentage of educated individuals in a community may influence HIV testing uptake (Alem, Liyew, & Guadie, 2021). In this study, similar factors may have led to low testing uptake, including differences in HIV knowledge, cultural differences, and different levels of stigma associated with HIV.

This study found that, in the home-based HIV testing conducted in Siaya between May 2016 and July 2017, 14% of clients who were eligible for testing were not tested; although the overall testing uptake was high (86%), and was comparable to the testing uptake (64% to 99%) reported in other home-based testing programs in sub-Saharan Africa (Sabapathy et al., 2012). Of clients not tested, the majority (75%) were not found at home at the time of testing, and a smaller proportion (14%) declined testing.

This study found that males compared to females; clients aged 20-24, 25-35 and >35 years compared to 15-19 years; and household heads, children and other relatives compared to spouses, had lower HIV testing uptake. Generally, among male and female, older age-groups

tended to have lower testing uptake. Previous studies have reported factors associated with lower home-based HIV testing uptake, and include: age>25years and having concurrent partnership at time of home-based testing (Helleringer et al., 2009); female counsellors approaching male clients, and wife of head of household man who's not tested or head of household being non-husband (Helleringer et al., 2009). On the other hand, factors reported in literature to be associated with higher home-based HIV testing uptake include: men (Sekandi et al., 2011); women (Dalal et al., 2013; Lugada et al., 2010; Tumwesigye, Wana, Kasasa, Muganzi, & Nuwaha, 2010); those never married, farmers, or older (>45years) head of household (Kranzer et al., 2008); previously married compared to never married (Sekandi et al., 2011); income bottom quartile (Helleringer et al., 2009); >35years when compared with 15-24 years (Lugada et al., 2010; Sekandi et al., 2011); currently married or divorced/widowed/separated compared to never married; and previously tested within 12 months compared to those never tested (Sekandi et al., 2011).

The biggest contributor to low home-based testing uptake in this study was clients not being found at home. Household heads, children, males≥25 years and females aged 15-19 years were more likely not found at home. This is due to older males (who are also likely the household heads) being more likely to go for work outside the home compared to females who are more likely to remain at home to do household chores, while younger age groups are likely to be away in school. Other studies have documented similar findings of older males being less likely to be found at home (Kranzer et al., 2008); while females are more likely to be found at home (Tumwesigye et al., 2010). Clients declining testing also contributed to low testing uptake. Clients who declined HIV testing in this study were more likely household heads and older (aged ≥25 years) males and females. Although this study did not document client reasons for declining

HIV testing, other studies have shown that not being ready/feeling scared/needing to think about it, fear of knowing one's status, not feeling at risk of having or acquiring HIV, preferring to test away from home, and wanting to test later, are reasons for declining home-based testing (Dalal et al., 2013; Kranzer et al., 2008; Naik et al., 2012a).

To increase HIV testing uptake, and ensure optimal identification of HIV positive clients, additional HIV testing strategies need to be developed tailored for clients who are missed by routinely offered facility or community testing strategies. From this study, testing strategies tailored for clients not found at home and those who declined testing should be developed; examples include mobile out-reaches for men at their workplace; testing in schools or colleges to reach younger men and women; enhanced pre-test counseling sessions for older men, women, and household heads, and social network testing mobilized by peers to increase testing acceptance. Furthermore, this study demonstrates that testing strategies tailored for clients with low testing uptake (including those not found at home and those who declined testing) can be further targeted to granular geographic areas with high HIV positive yield for efficiency, instead of being generalized.

The study limitations are similar to those described in study objective 2 above.

# CHAPTER SIX
# SUMMARY OF FINDINGS, CONCLUSIONS, RECOMMENDATIONS

## 6.1 Summary of findings

The following is a summary of the findings from the three study objectives:

1. A HIV predictive risk-score algorithm that was developed from a set of socio-demographic and behavioral characteristics, was successfully used to identify outpatient sub-populations that have higher risk of HIV infection, to whom HIV testing should be targeted. The overall final risk-score algorithm contained the following characteristics: age category 35–39/40–44 years; occupation (manual/domestic or trade/sales/service); marital status (polygamous marriage, separated/divorced or widowed); in the prior 12 months having ≥2 sexual partners or reporting treatment for an STI; and having never been tested for HIV or having a negative HIV test result >12 months prior. The overall algorithm's ability to discriminate between individuals with, and without, HIV infection in the general outpatient setting was modest. Additionally, using the three highest risk-score categories in the overall algorithm (risk score ≥10) to target HIV testing would dramatically reduce (by about 75%) the number of patients tested; however, it would miss the diagnosis of approximately 50% of HIV infected individuals accessing health facilities, making the use of the algorithm inferior to universal testing. In settings where universal testing is not feasible, the HIV predictive risk-score screening algorithm offers an evidence-base to guide identification of patient sub-populations with higher HIV risk, to whom HIV testing should be targeted.
2. Geospatial analysis was successfully used to identify granular-geographic clusters of new HIV diagnoses where HIV testing and other HIV interventions should be targeted.

Additionally, the study identified sub-populations with higher HIV positive yield (i.e., older age groups, those in polygamous marriage or separated divorced, and those never tested for HIV, or tested HIV-negative >12 months prior), that would benefit from continued targeted HIV testing and prevention interventions.

3. Mapping of HIV testing uptake successfully identified granular-geographic areas with low HIV testing uptake yet higher HIV positive yield, to inform targeting of HIV testing. Although the overall HIV testing uptake was high (86%), 14% of eligible clients were not tested; majority (75%) of whom were not found at home at the time of testing, and a smaller proportion (14%) who declined testing. This study identified granular-geographic areas with higher new HIV diagnoses and yet low testing uptake, high proportion of clients not found at home, and high proportion of clients who declined testing, that should be targeted to efficiently increase HIV testing uptake. Additionally, this study identified sub-populations with low HIV testing uptake (i.e. older age groups, males, household head, children and other relatives), those more likely not to be found at home (i.e. household heads, children, males>25 years of age, and females 15-19 years) and those more likely to decline testing (i.e. household heads, older males and females>25 years of age), that programs need to tailor HIV testing strategies to increase testing uptake.

### 6.2 Conclusion

This study demonstrates key strategies to identify sub-populations and granular-geographic areas with higher HIV positive yield, to inform strategies to target HIV testing among persons $\geq$15 years of age, for maximal epidemiologic impact and efficient use of resources:

1. A HIV predictive risk-score algorithm, derived from a set of socio-demographic and behavioral characteristics, that successfully identified sub-populations who have higher risk of HIV infection to whom HIV testing should be targeted.
2. Geospatial analysis in a routine public health program, that was successfully used to identify geographic clusters of higher new HIV diagnoses and sub-populations with higher HIV positive yield, that HIV testing and other HIV interventions should be targeted with finer granularity.
3. Mapping of HIV testing uptake, that successfully identified granular-geographic areas with low HIV testing uptake yet higher HIV positive yield, that HIV testing should be targeted.

This study's findings are important, as strategies to inform targeting of HIV testing would lead to efficient identification of HIV positive individuals, in order to link them to ART and reduce new HIV infections, and HIV-related morbidity and mortality. Significantly reducing new infections is critical in controlling the HIV epidemic, in order to realize both a public health (a significant reduction in morbidity and mortality) and economic (future significant cost-savings) impact globally.

### 6.3 Recommendations from the Current Study

1. The HIV predictive risk-score screening algorithm should be used by programs, in situations where universal testing is not feasible due to limited resources or logistical challenges; this includes many sub-Sahara Africa settings. The algorithm offers an evidence-base to guide identification of patient sub-populations with higher HIV risk, to whom HIV testing should be targeted.

2. Geospatial analysis of new diagnoses should be integrated into routine public health programs, to help focus interventions to more granular geographic units and sub-populations for maximal epidemiologic impact and efficient resource allocation.
3. Programs should conduct mapping of HIV testing uptake using routine data in order to identify granular geographic areas with low HIV testing uptake yet higher HIV positive yield to inform targeting of HIV testing.

### 6.4 Recommendations for future studies

1. For the HIV predictive risk-score screening algorithm, the following is recommended:
    - Since the performance of the algorithm was modest (AUC of 0.69), it is recommended that studies to improve the algorithm's performance be explored. These studies should include some of the predictor variables that were not included in this study, e.g., education level, having a HIV positive sexual partner, etc.
    - Further evaluation is needed to externally validate the HIV predictive risk-score screening algorithm in other settings outside of the western region of Kenya, and to assess the impact of its use.
2. For geospatial analysis and mapping, it is recommended that:
    - Since geospatial analysis and mapping are not used routinely in programs, an evaluation is done to assess the acceptance, use and impact, once it is introduced and staff are trained to use it.

## REFERENCES


Adeokun, L., & Nalwadda, R. (1997). Serial marriages and AIDS in Masaka District. *Health Transit Rev, 7*, 49-66. Retrieved from http://www.jstor.org/stable/40652292

Alem, A., Liyew, A., & Guadie, H. (2021). Spatial pattern and associated factors of HIV testing and counselling among youths (15–24 years) in Ethiopia. *BMC Public Health, 21*(1), 644. doi:10.1186/s12889-021-10677-0

Alere Medical Co. Ltd. (2015). Determine$^{TM}$ HIV 1/2 Package Insert. Retrieved from File:///C:/Users/hoz1/Downloads/120001935%20v01%20Alere%20Determine%20HIV-12%20Package%20Insert%20.pdf

Allison, W., Kiromat, M., Vince, J., Handan, C., Graham, S., & Kaldor, J. (2011). Development of a clinical algorithm to prioritise HIV testing of hospitalised paediatric patients in a low resource moderate prevalence setting. *Archives of Disease in Childhood, 96*(1), 67-72.

Amornkul, P., Vandenhoudt, H., Nasokho, P., Odhiambo, F., Mwaengo, D., Hightower, A., . . . Vitek, C. (2009). HIV Prevalence and Associated Risk Factors Among Individuals Aged 13-34 years in Rural Western Kenya. *PLOS One, 4*(7), e6470.

Anderson, S., Cherutich, P., Kilonzo, N., Cremin, I., Fecht, D., Kimanga, D., . . . Maina, W. (2014). Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: a modelling study. *The Lancet, 384*(9939), 249-256.

Annie, M. (2011). Spousal Intimate Partner Violence is Associated with HIV and Other STIs Among Married Rwandan Women. *AIDS Behav 15*(142). Retrieved from https://doi.org/10.1007/s10461-009-9526-1

Anselin, L., Ibnu, S., & Youngihn, K. (2016). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis.* Retrieved from https://geodacenter.github.io/

Anselin, L., Syabri, I., & Kho, Y. (2010). GeoDa: An Introduction to Spatial Data Analysis. In M. M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (pp. 73-89). Berlin, Heidelberg: Springer Berlin Heidelberg.

Auvert, B., Taljaard, D., Lagarde, E., Sobngwi-Tambekou, J., Sitta, R., & Puren, A. (2005). Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial. *PLOS Medicine, 2*(11), e298. doi:10.1371/journal.pmed.0020298

Ayisi, J., Van-Eijk, A., Kuile, F., Kolczak, M., Otieno, J., & Misore, A. (2000). Risk factors for HIV infection among asymptomatic pregnant women attending an antenatal clinic in western Kenya. *International Journal of STD & AIDS, 11*(6), 393-401.

Bailey, R., Moses, S., Parker, C., Agot, K., Maclean, I., Krieger, J., . . . Jeckoniah, O. (2007). Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *The Lancet, 369*, 643–656. doi:10.1016/S0140-6736(07)60312-2

Balkus, J., Brown, E., Palanee, T., Nair, G., Gafoor, Z., Zhang, J., . . . Baeten, J. (2016). An empiric HIV risk scoring tool to predict HIV-1 acquisition in African women. *Journal of Acquired Immune Deficiency Syndromes, 72*(3), 333.

Bandason, T., Dauya, E., Dakshina, S., McHugh, G., Chonzi, P., Munyati, S., . . . Ferrand, R. (2018). Screening tool to identify adolescents living with HIV in a community setting in Zimbabwe: A validation study. *PLOS One, 13*(10), e0204891. doi:10.1371/journal.pone.0204891

Bandasona, T., McHugha, G., Dauyaa, E., Mungofab, S., Munyatia, S., Weissc, H., . . . Ferrand, R. (2016). Validation of a screening tool to identify older children living with HIV in primary care facilities in high HIV prevalence settings. *AIDS*(30), 779–785.

Bandyopadhyay, A., Bhattacharyya, S., & Banerjee, A. (2009). Clinicoepidemiological scoring system for early diagnosis of pediatric HIV. *Indian Pediatrics, 46*(6), 512-515.

Bassett, I., Regan, S., Mbonambi, H., Blossom, J., Bogan, S., Bearnot, B., . . . Freedberg, K. (2015). Finding HIV in hard to reach populations: mobile HIV testing and geospatial mapping in Umlazi township, Durban, South Africa. *AIDS and Behavior, 19*(10), 1888-1895.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics, 43*(1), 1-20.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology, 4*, 33-49.

Boileau, C., Clark, S., Bignami-Van Assche, S., Poulin, M., Reniers, G., Watkins, S., . . . Heymann, S. (2009). Sexual and marital trajectories and HIV infection among ever-married women in rural Malawi. *Sexually Transmitted Infections, 85*(Suppl 1), i27-i33.

Bove, R., & Valeggia, C. (2009). Polygyny and women's health in sub-Saharan Africa. *Social Science & Medicine, 68*(1), 21-29.

Bulstra, C., Hontelez, J., Giardina, F., Steen, R., Nagelkerke, N., Bärnighausen, T., & de Vlas, S. (2020). Mapping and characterising areas with high levels of HIV transmission in sub-Saharan Africa: A geospatial analysis of national survey data. *PLOS Medicine, 17*(3), e1003042. doi:10.1371/journal.pmed.1003042

Busgeeth. (2004). The use of a spatial information system in the management of HIV/AIDS in South Africa *International Journal of Health Geographics, 3*(13). Retrieved from https://doi.org/10.1186/1476-072X-3-13

Centers for Disease Control and Prevention. (2023). The State of the HIV Epidemic in the U.S. Retrieved from https://www.cdc.gov/nchhstp/newsroom/fact-sheets/hiv/state-of-the-hiv-epidemic-factsheet.html

Chan, L., McGarey, P., & Sclafani, J. A. (2018). Using large data sets for population-based health research. In *Principles and practice of clinical research* (pp. 293-302): Elsevier.

Chen, Z., Branson, B., Ballenger, A., & Peterman, T. (1998). Risk assessment to improve targeting of HIV counseling and testing services for STD clinic patients. *Sexually Transmitted Diseases, 25*(10), 539-543.

Choudhry, V., Ambresin, A., Nyakato, V., & Agardh, A. (2015). Transactional sex and HIV risks–evidence from a cross-sectional national survey among young people in Uganda. *Global health action, 8*(1), 27249.

Clemens, S., Macneal, K., Alons, C., & Cohn, J. (2020). Screening Algorithms to Reduce Burden of Pediatric HIV Testing: A Systematic Review and Meta-analysis. *The Pediatric Infectious Disease Journal, 39*(10), e303-e309. doi:10.1097/inf.0000000000002715

Cohen, M., Chen, Y., McCauley, M., Gamble, T., Hosseinipour, M., Kumarasamy, N., . . . Pilotto, J. (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine, 365*(6), 493-505.

Cohen, M., Chen, Y., McCauley, M., Gamble, T., Hosseinipour, M., Kumarasamy, N., . . . Pilotto, J. (2016). Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *New England Journal of Medicine, 375*(9), 830–839. doi:doi:10.1056

Coombs, R., Reichelderfer, P., & Landay, A. (2003). Recent observations on HIV type-1 infection in the genital tract of men and women. *AIDS, 17*(4), 455-480.

Cuadros, D., Awad, S., & Abu-Raddad, L. (2013a). Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. *International Journal of Health Geographics, 12*(1), 28.

Cuadros, D., Awad, S., & Abu-Raddad, L. (2013b). Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. *International Journal of Health Geographics, 12*, 28. doi:10.1186/1476-072x-12-28

Cuadros, D., de Oliveira, T., Gräf, T., Junqueira, D., Wilkinson, E., Lemey, P., . . . Tanser, F. (2022). The role of high-risk geographies in the perpetuation of the HIV epidemic in rural South Africa: A spatial molecular epidemiology study. *PLOS Global Public Health, 2*(2), e0000105. doi:10.1371/journal.pgph.0000105

Dalal, W., Feikin, D., Amolloh, M., Ransom, R., Burke, H., Lugalia, F., . . . Breiman, R. (2013). Home-based HIV testing and counseling in rural and urban Kenyan communities. *Journal of Acquired Immune Deficiency Syndromes, 62*(2), e47-e54.

Debray, T., Vergouwe, Y., Koffijberg, H., Nieboer, D., Steyerberg, E., & Moons, K. (2015). A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of Clinical Epidemiology, 68*(3), 279-289.

Dieffenbach, C. (2012). Preventing HIV transmission through antiretroviral treatment-mediated virologic suppression: aspects of an emerging scientific agenda. *Current Opinion in HIV and AIDS, 7*, 106–110.

Dube, B., Marshall, T., Ryan, R., & Omonijo, M. (2018). Predictors of human immunodeficiency virus (HIV) infection in primary care among adults living in developed countries: a systematic review. *Systematic Reviews, 7*(1), 82.

Dude, A. (2011). Spousal Intimate Partner Violence is Associated with HIV and Other STIs Among Married Rwandan Women. *AIDS and Behavior, 15*(142). Retrieved from https://doi.org/10.1007/s10461-009-9526-1

Dyer, J., Eron, J., Hoffman, I., Kazembe, P., Vernazza, P., Nkata, E., . . . Cohen, M. (1998). Association of CD4 cell depletion and elevated blood and seminal plasma human immunodeficiency virus type 1 (HIV-1) RNA concentrations with genital ulcer disease in HIV-1-infected men in Malawi. *The Journal of Infectious Diseases, 177*(1), 224-227.

Eberhart, M., Yehia, B., Hillier, A., Voytek, C., Blank, M., Frank, I., . . . Brady, K. (2013). Behind the Cascade: Analyzing Spatial Patterns along the HIV Care Continuum. *Journal*

*of Acquired Immune Deficiency Syndromes, 64*(01), S42–S51. doi: 10.1097/QAI.0b013e3182a90112

Eccles, M. P., & Mittman, B. S. (2006). Welcome to Implementation Science. *Implementation Science, 1*(1), 1. doi:10.1186/1748-5908-1-1

Elías, M., Gómez-Ayerbe, C., Elías, P., Muriel, A., de Alberto, S., Martinez-Colubi, M., . . . Barea, R. (2016). Development and validation of an HIV risk exposure and indicator conditions questionnaire to support targeted HIV screening. *Medicine, 95*(5), e2612.

Farmer, P. (2001). *Infections and inequalities: The modern plagues*: Univ of California Press.

Ferguson, G., & Morris, C. N. (2007). Mapping transactional sex on the Northern Corridor highway in Kenya *Health & Place, 13*(2), 504-519. Retrieved from https://doi.org/10.1016/j.healthplace.2006.05.009

Ferrand, R., Weiss, H., Nathoo, K., Ndhlovu, C., Mungofa, S., Munyati, S., . . . Corbett, E. (2011). A primary care level algorithm for identifying HIV-infected adolescents in populations at high risk through mother-to-child transmission. *Tropical Medicine & International Health, 30*(3), 349–355.

Fleming, D., & Wasserheit, J. (1999). From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sexually Transmitted Infections, 75*(1), 3-17.

Gerbert, B., Bronstone, A., McPhee, S., Pantilat, S., & Allerton, M. (1998). Development and testing of an HIV-risk screening instrument for use in health care settings. *American Journal of Preventive Medicine, 15*(2), 103-113.

Ghys, P., Fransen, K., Diallo, M., Ettiègne-Traoré, V., Coulibaly, I., Yeboué, K., . . . Greenberg, A. (1997). The associations between cervicovaginal HIV shedding, sexually transmitted diseases and immunosuppression in female sex workers in Abidjan, Cote d'Ivoire. *AIDS, 11*(12), F85-F93.

Giovenco, D., Pettifor, A., MacPhail, C., Kahn, K., Wagner, R., Piwowar-Manning, E., . . . Hughes, J. (2019). Assessing risk for HIV infection among adolescent girls in South Africa: an evaluation of the VOICE risk score (HPTN 068). *Journal of the International AIDS Society, 22*(7), e25359.

González, R., Augusto, O., Munguambe, K., Pierrat, C., Pedro, E., Sacoor, C., . . . Alonso, P. (2015). HIV incidence and spatial clustering in a rural area of Southern Mozambique. *PLOS One, 10*(7), e0132053.

Grabowski, M. K., Burke, V., Nakiigozi, G., Nalugoda, F., Ssekubugu, R., & Chang, L. (2018). *Transactional sex measurement and association with HIV incidence among women.* Paper presented at the Conference on Retroviruses and Opportunistic Infections (CROI).

Group., I. S. S. (2015). Initiation of antiretroviral therapy in early asymptomatic HIV infection. *New England Journal of Medicine, 373*(9), 795-807.

Hargreaves, J., & Glynn, J. (2002). Educational attainment and HIV-1 infection in developing countries: a systematic review. *Tropical Medicine & International Health, 7*(6), 489-498.

Hargreaves, J., Morison, L., Chege, J., Rutenburg, N., Kahindo, M., Weiss, H., . . . Buvé, A. (2002). Socioeconomic status and risk of HIV infection in an urban population in Kenya. *Tropical Medicine & International Health, 7*(9), 793-802.

Harrell, F., Frank, E., Lee, K., & Mark, D. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine, 15*(4), 361-387.

Haukoos, J., Hopkins, E., Bender, B., Sasson, C., Al-Tayyib, A., & Thrun, M. (2013). Comparison of enhanced targeted rapid HIV screening using the Denver HIV risk score to nontargeted rapid HIV screening in the emergency department. *Annals of Emergency Medicine, 61*(3), 353-361.

Haukoos, J., Hopkins, E., Bucossi, M., Lyons, M., Rothman, R., White, D., . . . Sabel, A. (2015). Validation of a quantitative HIV risk prediction tool using a national HIV testing cohort. *Journal of Acquired Immune Deficiency Syndromes, 68*(5), 599.

Haukoos, J., Lyons, M., Lindsell, C., Hopkins, E., Bender, B., Rothman, R., . . . Sasson, C. (2012). Derivation and validation of the Denver Human Immunodeficiency Virus (HIV) risk score for targeted HIV screening. *American Journal of Epidemiology, 175*(8), 838-846.

Helleringer, S., Kohler, H., Frimpong, J., & Mkandawire, J. (2009). Increasing uptake of HIV testing and counseling among the poorest in sub-Saharan countries through home-based service provision. *Journal of Acquired Immune Deficiency Syndromes, 51*(2), 185.

Hixson, B., Omer, S., Del Rio, C., & Frew, P. (2011). Spatial Clustering of HIV Prevalence in Atlanta, Georgia and Population Characteristics Associated with Case Concentrations. *Journal of Urban Health, 88*(1), 129-141.

Hoenigl, M., Weibel, N., Mehta, S., Anderson, C., Jenks, J., Green, N., . . . Little, S. (2015). Development and validation of the San Diego Early Test Score to predict acute and early HIV infection risk in men who have sex with men. *Clinical Infectious Diseases, 61*(3), 468-475.

Horwood, C., Liebeschuetz, S., Blaauw, D., Cassol, S., & Qazi, S. (2003). Diagnosis of paediatric HIV infection in a primary health care setting with a clinical algorithm. *Bulletin of the World Health Organization, 81*, 858-866.

Horwood, C., Vermaak, K., Rollins, N., Haskins, L., Nkosi, P., & Qazi, S. (2009). Paediatric HIV management at primary care level: an evaluation of the integrated management of childhood illness (IMCI) guidelines for HIV. *BMC Pediatrics, 9*, 59. doi:10.1186/1471-2431-9-59

Hsieh, Y., Haukoos, J., & Rothman, R. (2014). Validation of an abbreviated version of the Denver HIV risk score for prediction of HIV infection in an urban ED. *American Journal of Emergency Medicine, 32*(7), 775-779.

Johnson, L. (2018). Design of observational studies. In *Principles and practice of clinical research* (pp. 231-248): Elsevier.

Joint United Nations Programme on HIV/AIDS. (2014). *90-90-90: An ambitious treatment target to help end the AIDS epidemic*. Retrieved from https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf

Joint United Nations Programme on HIV/AIDS. (2020a). 2025 AIDS Targets. Retrieved from https://aidstargets2025.unaids.org/

Joint United Nations Programme on HIV/AIDS. (2020b). *UNAIDS DATA 2022*. Retrieved from https://www.unaids.org/sites/default/files/media_asset/data-book-2022_en.pdf

Joint United Nations Programme on HIV/AIDS. (2020c). UNAIDS Facts Sheet- World AIDS Day 2020. Retrieved from https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf

Joint United Nations Programme on HIV/AIDS. (2021). *UNAIDS data 2021*. Retrieved from https://www.unaids.org/en/resources/documents/2021/2021_unaids_data

Jones, J., Hoenigl, M., Siegler, A., Sullivan, P., Little, S., & Rosenberg, E. (2017). Assessing the performance of 3 human immunodeficiency virus incidence risk scores in a cohort of black and white men who have sex with men in the South. *Sexually Transmitted Diseases, 44*(5), 297-302.

Joseph, R., Musingila, P., Miruka, F., Wanjohi, S., Dande, C., Musee, P., . . . Okomo, G. (2019). Expanded eligibility for HIV testing increases HIV diagnoses—A cross-sectional study in seven health facilities in western Kenya. *PLOS One, 14*(12).

Kalichman, S., Simbayi, L., Jooste, S., Vermaak, R., & Cain, D. (2008). Sensation seeking and alcohol use predict HIV transmission risks: prospective study of sexually transmitted infection clinic patients, Cape Town, South Africa. *Addictive Behaviors, 33*(12), 1630-1633.

Kansiime, S., Hansen, C., Hayes, R., & Ruzagira, E. (2023). Developing HIV risk prediction tools in four African settings. *Tropical Medicine & International Health, 28*(9), 720-730. doi:10.1111/tmi.13916

Kapiga, S., Lyamuya, E., Vuylsteke, B., Spiegelman, D., Larsen, U., & Hunter, D. (2000). Risk factors for HIV-1 seroprevalence among family planning clients in Dar es Salaam, Tanzania. *African Journal of Reproductive Health, 4*(1), 88-99.

Karim, S., Richardson, B., Ramjee, G., Hoffman, I., Chirenje, Z., Taha, T., . . . Profy, A. (2011). Safety and effectiveness of BufferGel and 0.5% PRO2000 gel for the prevention of HIV infection in women. *AIDS, 25*(7), 957-966.

Kenya Ministry of Health. (2015). *Kenya HIV Testing Guidelines* Retrieved from https://aidsfree.usaid.gov/sites/default/files/hts_policy_kenya_2015.pdf

Kenya National AIDS and STI Control Programme. (2020). *Kenya HIV Estimates*. Retrieved from https://nsdcc.go.ke/download/kenya-hiv-estimates-report-2020/

Kenya National AIDS Control Council. (2009). *Kenya HIV Prevention Response and Modes of Transmission Analysis* Retrieved from https://icop.or.ke/wp-content/uploads/2016/09/KenyaMOT-2009.pdf

Kenya National Bureau of Statistics. (2009). Kenya Population and Housing Census. Retrieved from file:///C:/Users/hoz1/Downloads/Population%20Distribution%20by%20Sex,%20Number

%20of%20Households,%20Area%20and%20Density%20by%20County%20and%20District.pdf

Kenya National Bureau of Statistics. (2014). Kenya Demographic and Health Survey 2014. Retrieved from https://dhsprogram.com/pubs/pdf/fr308/fr308.pdf

Kenya National Bureau of Statistics. (2019). Kenya Population and Housing Census Retrieved from https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county

Kenya State Department for Devolution. (n.d). County Information. Retrieved from https://www.devolution.go.ke/county-information/

Kharsany, A., & Karim, Q. (2016). HIV infection and AIDS in sub-Saharan Africa: current status, challenges and opportunities. *The Open AIDS Journal, 10*, 34.

Kim, A., Mukui, I., Young, P., Mirjahangir, J., Mwanyumba, S., Wamicwe, J., . . . De Cock, K. (2016). Undisclosed HIV infection and antiretroviral therapy use in the Kenya AIDS indicator survey 2012: relevance to national targets for HIV diagnosis and treatment. *AIDS, 30*(17), 2685-2695. doi:https://doi.org/10.1097/QAD. 0000000000001227

Kim, H., Musuka, G., Mukandavire, Z., Branscum, A., & Cuadros, D. (2021). When distance matters: Mapping HIV health care underserved communities in sub-Saharan Africa. *PLOS Global Public Health, 1*(11), e0000013. doi:10.1371/journal.pgph.0000013

Kimanga, D., Ogola, S., & Umuro, M. (2014). Prevalence and incidence of HIV infection, trends, and risk factors among persons aged 15–64 years in Kenya: results from a nationally representative study. *Journal of Acquired Immune Deficiency Syndromes, 66*(Suppl 1), S13.

Kimani, J., Ettarh, R., Ziraba, A., & Yatich, N. (2011). Marital status and risk of HIV infection in informal urban settlements of Nairobi, Kenya: results from a cross-sectional survey. *Journal of Epidemiology & Community Health, 65*(Suppl 1), A340-A341.

Kimani, J., Ettarh, R., Ziraba, A., & Yatich, N. (2013). Marital status and risk of HIV infection in slum settlements of Nairobi, Kenya: results from a cross-sectional survey. *African Journal of Reproductive Health, 17*(1), 103-113.

Kish, L. (1979). Samples and censuses. *International Statistical Review/Revue Internationale de Statistique*, 99-109.

Kranzer, K., Saul, J., Crampin, A., Jahn, A., Malema, S., Mulawa, D., . . . Glynn, J. (2008). Individual, household and community factors associated with HIV test refusal in rural Malawi. *Tropical Medicine & International Health, 13*(11), 1341-1350.

Kulldorff, M., & Nagarwalla, N. (1995). Spatial disease clusters: detection and inference. *Statistics in Medicine, 14*(8), 799-810.

Laupacis, A., & Sekar, N. (1997). Clinical prediction rules: a review and suggested modifications of methodological standards. *JAMA, 277*(6), 488-494.

Lazzarin, A., Saracco, A., Musicco, M., & Nicolosi, A. (1991). Man-to-woman sexual transmission of the human immunodeficiency virus: risk factors related to sexual behavior, man's infectiousness, and woman's susceptibility. *Archives of Internal Medicine, 151*(12), 2411-2416.

Lilian, R., Grobbelaar, C., Hurter, T., McIntyre, J., Struthers, H., & Peters, R. (2017). Application opportunities of geographic information systems analysis to support achievement of the UNAIDS 90-90-90 targets in South Africa. *South African Medical Journal, 107*(12), 1065-1071.

Lugada, E., Levin, J., Mermin, J., Mugalanzi, E., Namara, G., Gupta, S., . . . Coutinho, A. (2010). Comparison of home and clinic-based HIV testing among household members of persons taking antiretroviral therapy in Uganda: results from a randomized trial. *Journal of Acquired Immune Deficiency Syndromes, 55*(2), 245-252.

Lyons, M., Lindsell, C., Ruffner, A., Wayne, D., Hart, K., Sperling, M., . . . Fichtenbaum, C. (2013). Randomized comparison of universal and targeted HIV screening in the emergency department. *Journal of Acquired Immune Deficiency Syndromes, 64*(3), 315.

Marrazzo, J., Ramjee, G., Richardson, B., Gomez, K., Mgodi, N., Nair, G., . . . Noguchi, L. (2015). Tenofovir-based preexposure prophylaxis for HIV infection among African women. *New England Journal of Medicine, 372*(6), 509-518.

Martin Kulldorff and Information Management Services Inc. (2009, 19 October 2019). SaTScan™: Software for the spatial and space-time scan statistics. Retrieved from http://www.satscan.org/. Accessed 3 October 2020.

Martin Kulldorff and Information Management Services Inc. (2018). SaTScanTM User Guide for version 9.6. Retrieved from https://www.satscan.org/cgi-

bin/satscan/register.pl/SaTScan_Users_Guide.pdf?todo=process_userguide_download. Accessed 7 October 2020.

Mbopi-Kéou, F., Grésenguet, G., Mayaud, P., Weiss, H., Gopal, R., Matta, M., . . . Mabey, D. (2000). Interactions between herpes simplex virus type 2 and human immunodeficiency virus type 1 infection in African women: opportunities for intervention. *The Journal of Infectious Diseases, 182*(4), 1090-1096.

Menza, T., Hughes, J., Celum, C., & Golden, M. (2009). Prediction of HIV acquisition among men who have sex with men. *Sexually Transmitted Diseases, 36*(9), 547.

Ministry of Health National AIDS & STI Control Program. (2022). *Kenya HIV Prevention and Treatment Guidelines*. Retrieved from https://www.differentiatedservicedelivery.org/wp-content/uploads/Kenya-ARV-Guidelines-2022-Final-1.pdf

Mollié, A. ( 1996). *Bayesian mapping of disease.* (Vol. 1).

Moons, K., Altman, D., Reitsma, J., Ioannidis, J., Macaskill, P., Steyerberg, E., . . . Collins, G. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine, 162*(1), W1-W73.

Moucheraud, C., Chasweka, D., Nyirenda, M., Schooley, A., Dovel, K., & Hoffman, R. (2018). Simple Screening Tool to Help Identify High-Risk Children for Targeted HIV Testing in Malawian Inpatient Wards. *Journal of Acquired Immune Deficiency Syndromes, 79*(3), 352-357. doi:10.1097/qai.0000000000001804

Moucheraud, C., Hoffman, R., Balakasi, K., Wong, V., Sanena, M., Gupta, S., & Dovel, K. (2022). Screening Adults for HIV Testing in the Outpatient Department: An Assessment of Tool Performance in Malawi. *AIDS and Behavior, 26*(2), 478-486. doi:10.1007/s10461-021-03404-8

Msisha, W., Kapiga, S., Earls, F., & Subramanian, S. (2008). Socioeconomic status and HIV seroprevalence in Tanzania: a counterintuitive relationship. *International Journal of Epidemiology, 37*(6), 1297-1303. doi:10.1093/ije/dyn186

Muloongo, K., Tshuma, N., Chimoyi, L., Setswe, G., Sarfo, B., & Nyasulu, P. (2014). Factors contributing to home-based acceptability of rapid testing for HIV infection among the inner city commuter population in Johannesburg, South Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 108*(10), 632-638. doi:10.1093/trstmh/tru122

Muttai, H., Guyah, B., Achia, T., Musingila, P., Nakhumwa, J., Oyoo, R., . . . Agot, K. (2021). Mapping geographic clusters of new HIV diagnoses to inform granular-level interventions for HIV epidemic control in western Kenya. *BMC Public Health, 21*(1), 1-15.

Muttai, H., Guyah, B., Musingila, P., Achia, T., Miruka, F., Wanjohi, S., . . . Onyango, D. (2021). Development and validation of a sociodemographic and behavioral characteristics-based risk-score algorithm for targeting HIV testing among adults in Kenya. *AIDS and Behavior, 25*(2), 297-310.

Naik, R., Tabana, H., Doherty, T., Zembe, W., & Jackson, D. (2012a). Client characteristics and acceptability of a home-based HIV counselling and testing intervention in rural South Africa. *BMC Public Health, 12*(1), 824. doi:10.1186/1471-2458-12-824

Naik, R., Tabana, H., Doherty, T., Zembe, W., & Jackson, D. (2012b). Client characteristics and acceptability of a home-based HIV counselling and testing intervention in rural South Africa. *BMC Public Health, 12*, 824. doi:10.1186/1471-2458-12-824

Nalugoda, F., Guwatudde, D., Bwaninka, J., Makumbi, F., Lutalo, T., Kagaayi, J., . . . Gray, R. (2014). Marriage and the risk of incident HIV infection in Rakai, Uganda. *Journal of Acquired Immune Deficiency Syndromes, 65*(1), 91-98. doi:10.1097/QAI.0b013e3182a7f08a

National AIDS and STI Control Programme. (2008). *Kenya AIDS indicator survey 2012*. Retrieved from https://nsdcc.go.ke/wp-content/uploads/2015/10/KAIS-2012.pdf

National AIDS STI Control Programme Ministry of Health Kenya. (2013). Kenya AIDS indicator survey 2012. Retrieved from https://nacc.or.ke/wp-content/uploads/2015/10/KAIS-2012.pdf

Nilsen, P. (2020). *Making sense of implementation theories, models, and frameworks*.

Nutor, J., Duah, H., Duodu, P., Agbadi, P., Alhassan, R., & Darkwah, E. (2021). Geographical variations and factors associated with recent HIV testing prevalence in Ghana: spatial mapping and complex survey analyses of the 2014 demographic and health surveys. *BMJ Open, 11*(7), e045458. doi:10.1136/bmjopen-2020-045458

Nwachukwu, C., & Odimegwu, C. (2011). Regional patterns and correlates of HIV voluntary counselling and testing among youths in Nigeria. *African Journal of Reproductive Health, 15*(2), 131-146.

Oluoch, T., Mohammed, I., Bunnell, R., Kaiser, R., Kim, A., Gichangi, A., . . . Orago, A. (2011). Correlates of HIV infection among sexually active adults in Kenya: a national population-based survey. *The Open AIDS Journal, 5*, 125.

Ong, J., Coulthard, K., Quinn, C., Tang, M., Huynh, T., Jamil, M., . . . Johnson, C. (2021). Risk-Based Screening Tools to Optimise HIV Testing Services: A Systematic Review. *Current HIV/AIDS Reports*.

Patterson, B., Landay, A., Siegel, J., Flener, Z., Pessis, D., Chaviano, A., & Bailey, R. (2002). Susceptibility to human immunodeficiency virus-1 infection of human foreskin and cervical tissue grown in explant culture. *The American Journal of Pathology, 161*(3), 867-873.

Peduzzi, P., Concato, J., Kemper, E., Holford, T., & Feinstein, A. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology, 49*(12), 1373-1379.

Pettifor, A., Rees, H., Kleinschmidt, I., Steffenson, A., MacPhail, C., Hlongwa-Madikizela, L., . . . Padian, N. (2005). Young people's sexual health in South Africa: HIV prevalence and sexual behaviors from a nationally representative household survey. *AIDS, 19*(14), 1525-1534.

Porter, L., Hao, L., Bishai, D., Serwadda, D., Wawer, M., Lutalo, T., . . . Team., R. P. (2004). HIV status and union dissolution in sub-Saharan Africa: the case of Rakai, Uganda. *Demography, 41*(3), 465-482.

Premier Medical Corporation Limited. First Response®HIV 1-2-0 Human Immunodeficiency Virus Rapid Test Strip. Retrieved from http://premiermedcorp.com/wp-content/uploads/2017/09/I05FRS50-1.pdf. Accessed 2 October 2020.

President's Emergency Plan for AIDS Relief. (2020). *PEPFAR Annual Report*. Retrieved from https://pepfar-panorama.org

QGIS.org. QGIS Geographic Information System. Retrieved from http://qgis.org. Accessed 3 October 2020.

Quinn C., W. V., Jamil MS., et al. . (2020). *Web Annex L. Symptom and risk-based screening to optimize HIV testing services: a scoping review. In: Consolidated guidelines on HIV testing services, 2019*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/335903/9789240011830-eng.pdf

The R-INLA project, . Retrieved from http://www.r-inla.org/home. Accessed 16 February 2020. Retrieved 14 October 2019 http://www.r-inla.org/home. Accessed 16 February 2020.

R Core Team, R. (2013). R: A language and environment for statistical computing. Retrieved from http://www.R-project.org/

Ramspek, C., Jager, K., Dekker, F., Zoccali, C., & van Diepen, M. (2021). External validation of prognostic models: what, why, how, when and where? *Clinical Kidney Journal, 14*(1), 49-58.

Rivero, A., & Moreno, S. (2017). Is it time to start new HIV prevention strategies in Spain? *Enfermedades Infecciosas y Microbiología Clínica, 35*(5), 271-272.

Rosenberg, E., Delaney, K., Branson, B., Spaulding, A., Sullivan, P., & Sanchez, T. (2012). Re:"Derivation and validation of the Denver human immunodeficiency virus (HIV) risk score for targeted HIV screening". *American Journal of Epidemiology, 176*(6), 567-568.

Røttingen, J., Cameron, D., & Garnett, G. (2001). A systematic review of the epidemiologie interactions between classic sexually transmitted diseases and HIV: how much really is known? *Sexually Transmitted Diseases*, 579-597.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 71*(2), 319-392.

Ruzagira, E., Baisley, K., Kamali, A., & Grosskurth, H. (2018). Factors associated with uptake of home-based HIV counselling and testing and HIV care services among identified HIV-positive persons in Masaka, Uganda. *AIDS Care, 30*(7), 879-887. doi:10.1080/09540121.2018.1441967

Sabapathy, K., Van den Bergh, R., Fidler, S., Hayes, R., & Ford, N. (2012). Uptake of home-based voluntary HIV testing in sub-Saharan Africa: a systematic review and meta-analysis. *PLOS Medicine, 9*(12), e1001351.

Scott, H., Vittinghoff, E., Irvin, R., Liu, A., Nelson, L., Del Rio, C., . . . Van Tieu, H. (2020). Development and validation of the personalized sexual health promotion (SexPro) HIV risk prediction model for men who have sex with men in the United States. *AIDS and Behavior, 24*, 274-283.

Sekandi, J., Sempeera, H., List, J., Mugerwa, M., Asiimwe, S., Yin, X., & Whalen, C. (2011). High acceptance of home-based HIV counseling and testing in an urban community setting in Uganda. *BMC Public Health, 11*(1), 1-8.

Smith, D., Pals, S., Herbst, J., Shinde, S., & Carey, J. (2012). Development of a clinical screening index predictive of incident HIV infection among men who have sex with men in the United States. *Journal of Acquired Immune Deficiency Syndromes, 60*(4), 421-427.

Szabo, R., & Short, R. (2000). How does male circumcision protect against HIV infection? *The BMJ, 320*(7249), 1592-1594.

Talari, K., & Goyal, M. (2020). Retrospective studies–utility and caveats. *Journal of the Royal College of Physicians of Edinburgh, 50*(4), 398-402.

Tanser, F., Bärnighausen, T., Dobra, A., & Sartorius, B. (2017). Identifying 'corridors of HIV transmission' in a severely affected rural South African population: a case for a shift toward targeted prevention strategies. *International Journal of Epidemiology, 47*(2), 537-549.

Tenkorang, E. (2014). Marriage, widowhood, divorce and HIV risks among women in sub-Saharan Africa. *International Health, 6*(1), 46-53.

Toll, D., Janssen, K., Vergouwe, Y., & Moons, K. (2008). Validation, updating and impact of clinical prediction rules: a review. *Journal of Clinical Epidemiology, 61*(11), 1085-1094.

Tumwesigye, E., Wana, G., Kasasa, S., Muganzi, E., & Nuwaha, F. (2010). High uptake of home-based, district-wide, HIV counseling and testing in Uganda. *AIDS patient care and STDs, 24*(11), 735-741.

United States President's Emergency Plan for AIDS Relief. (2015). *PEPFAR Annual Report*. Retrieved from https://pepfar-panorama.org

United States President's Emergency Plan for AIDS Relief. (2016). *PEPFAR Annual Report*. Retrieved from https://pepfar-panorama.org

United States President's Emergency Plan for AIDS Relief. (2017). *PEPFAR Annual Report*. Retrieved from https://pepfar-panorama.org

US President's Emergency Plan for AIDS Relief. (2017). Strategy for Accelerating HIV/AIDS Epidemic Control (2017-2020). Retrieved from https://www.state.gov/wp-content/uploads/2019/08/PEPFAR-Strategy-for-Accelerating-HIVAIDS-Epidemic-Control-2017-2020.pdf

Van Damme, L., Corneli, A., Ahmed, K., Agot, K., Lombaard, J., Kapiga, S., . . . Onyango, J. (2012). Preexposure prophylaxis for HIV infection among African women. *New England Journal of Medicine, 367*(5), 411-422.

Wahome, E., Thiong'o, A., Mwashigadi, G., Chirro, O., Mohamed, K., Gichuru, E., . . . Sanders, E. (2018). An empiric risk score to guide PrEP targeting among MSM in coastal Kenya. *AIDS and Behavior, 22*, 35-44.

Wand, H., & Ramjee, G. (2010). Targeting the hotspots: investigating spatial and demographic variations in HIV infection in small communities in South Africa. *Journal of the International AIDS Society, 13*(1), 41.

Wand, H., Reddy, T., Naidoo, S., Moonsamy, S., Siva, S., Morar, N., & Ramjee, G. (2018). A simple risk prediction algorithm for HIV transmission: results from HIV prevention trials in KwaZulu Natal, South Africa (2002–2012). *AIDS and Behavior, 22*, 325-336.

Waruru, A., Achia, T., Tobias, J., J., N. a. a., Mwangi, M., Wamicwe, J., . . . Tylleskar, T. (2018). Finding Hidden HIV Clusters to Support Geographic-Oriented HIV Interventions in Kenya. *Journal of Acquired Immune Deficiency Syndromes, 78*(2), 144.

Waruru, A., Wamicwe, J., Mwangi, J., Achia, T., Zielinski-Gutierrez, E., Ng'ang'a, L., . . . Tylleskär, T. (2021a). Where Are the Newly Diagnosed HIV Positives in Kenya? Time to Consider Geo-Spatially Guided Targeting at a Finer Scale to Reach the "First 90". *Front Public Health, 9*, 503555. doi:10.3389/fpubh.2021.503555

Waruru, A., Wamicwe, J., Mwangi, J., Achia, T., Zielinski-Gutierrez, E., Ng'ang'a, L., . . . Tylleskär, T. (2021b). Where Are the Newly Diagnosed HIV Positives in Kenya? Time to Consider Geo-Spatially Guided Targeting at a Finer Scale to Reach the "First 90". *Front Public Health, 9*, 392.

Wasson, J., Sox, H., Neff, R., & Goldman, L. (1985). Clinical prediction rules: applications and methodological standards. *New England Journal of Medicine, 313*(13), 793-799.

Wawer, M., Makumbi, F., Kigozi, G., Serwadda, D., Watya, S., Nalugoda, F., . . . Moulton, L. (2009). Circumcision in HIV-infected men and its effect on HIV transmission to female partners in Rakai, Uganda: a randomised controlled trial. *The Lancet, 374* 229-237.

Wickham, H. (2016). ggplot2-Elegant Graphics for Data Analysis. Springer International Publishing. *Cham, Switzerland.* Retrieved from https://ggplot2.tidyverse.org

Winter, A., Taylor, S., Workman, J., White, D., Ross, J., Swan, A., & Pillay, D. (1999). Asymptomatic urethritis and detection of HIV-1 RNA in seminal plasma. *Sexually Transmitted Infections, 75*(4), 261-263.

Wojcicki, J. (2005). Socioeconomic status as a risk factor for HIV infection in women in East, Central and Southern Africa: a systematic review. *Journal of Biosocial Science, 37*(1), 1-36.

World Health Organization. (2010). *Preventing intimate partner and sexual violence against women: Taking action and generating evidence* (9241564008). Retrieved from https://www.who.int/publications/i/item/9789241564007

World Health Organization. (2015, 23 May 2018). Consolidated guidelines on HIV testing services. Retrieved from http://www.who.int/hiv/pub/guidelines/hiv-testing-services/en/

World Health Organization. (2016). Guidelines on HIV self-testing and partner notification: Supplement to consolidated guidelines on HIV testing services. Retrieved from http://www9.who.int/hiv/pub/self-testing/hiv-self-testing-guidelines/en/

World Health Organization. (2019). *WHO encourages countries to adapt HIV testing strategies in response to changing epidemic: policy brief*. Retrieved from https://www.who.int/publications/i/item/WHO-CDS-HIV-19.34

World Health Organization. (2021). *Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring: recommendations for a public health approach*. Retrieved from https://www.who.int/publications/i/item/9789240031593

Yin, L., Zhao, Y., Peratikos, M., Song, L., Zhang, X., Xin, R., . . . Hu, Y. (2018). Risk prediction score for HIV infection: development and internal validation with cross-sectional data from men who have sex with men in China. *AIDS and Behavior, 22*, 2267-2276.

Zablotska, I., Gray, R., Serwadda, D., Nalugoda, F., Kigozi, G., Sewankambo, N., . . . Wawer, M. (2006). Alcohol use before sex and HIV acquisition: a longitudinal study in Rakai, Uganda. *AIDS, 20:*, 1191–1196.

Zhu, Q., JiKe, C., Xu, C., Liang, S., Yu, G., Wang, J., . . . Jin, C. (2021). A New Strategy to Quantitatively Identify Hot-Spot Areas in Growth of New HIV Infections for Targeted Interventions. *Front Public Health, 9*, 680867. doi:10.3389/fpubh.2021.680867

# APPENDICES

## Appendix 1: HIV Behavioral Questionnaire Implemented in HIV Testing

### HIV Behavioral Risk Questionnaire/*Maswali ya Kutatua Tabia Hatari/ Penjo makonyowa ng'eyo timbe maketo ji e thuolo ma malo mar yudo kute mag ayaki*

**Facility Name/Site** *(Jina La Kituo/Nying kar thieth)* _____

**Ward Name** *(Jina la Kata/ Nying Kar thieth)* _____

**MFL code** *(Msimbo wa Kituo/ Namba kar thieth)* _____

**Date** *(Tarehe/ Tarik)* **(dd/mm/yyyy)** ____/____/_____

**Counselor's Name** *(Jina La Mshauri/ Nying jahocho)* _____

**HTS Client Code Number** *(Msimbo wa Mteja /Mshtiri/ Namba mar jaduoko)* ☐

**Age** *(Umri/ Higni)* _____**years**

**Sex** (Jinsia /kit chuech) ☐ **Male** (*Mme/N'gama dichuo)* ☐ **Female** (*Mke/ N'gama dhako)*

**1. Occupation** (*Kazi/Tich mitiyo)*_____

**2. Have you had sex in the last 12 months?** (*Je, umeshiriki ngono miezi 12 iliyopita? / Be isebedo e achiel e ringruok gi ng'ama dichuo kata dhako kuom dweche 12 mosekalo?)*

☐ Yes (*Ndio/ kamano)* ☐ No (*La/Ooyo)*

**3. Specify the number of sexual partner(s) you have had sex with in the last 12 months** (Tick one) (*Kama Ndiyo, umeshiriki ngono na wangapi miezi 12 iliyopita?/ Ka en kamano, isebediye achiel e ringruok kod ji adi kuom dweche apar gi ariyo mosekalo?)*

☐ Insert number (*Jumuisha idadi/ Gin ji adi*?) ☐

☐ Cannot remember the number, but ≤5 (*siwezi kumbuka lakini hawazidi tano/Ok anyal paro ni gin ji adi to ok ng'eny ne ji abich)*

☐ Cannot remember the number but >5 *(siwezi kumbuka lakini wamepita tano/ Ok anyal paro ni gin ji adi to okadho ji abich)*

☐ Don't know (*Sijui/Akia)*

☐ Not had sex in the last 12 months *(sijashiriki ngono kwa miezi 12 iliyopita/pok abedo e achiel kuom dweche 12 mokadho)*

☐ Decline to answer

**4. Have you had any change in your sexual partner(s) in the past 12 months?** /*Je, umewahi badilisha uhusiano wa kimapenzi kwa miezi 12 iliyopita/ Be ise bedo gi jahera machielo ma opogore gi mari ma pilepile kuom dweche 12 mosekadho?* ***(Tick all that apply)***

☐ Not had change in sexual partner in the last 12 months (*Sijawahi badilisha mpenzi kwa miezi 12 iliyopita/Pok aloko jaherana e dweche 12 mokadho)*

☐ New sexual partner(s) (*Nina mpenzi mpya/an gi jahera manyien)*

☐ Newly marriage in the last 12 months (*Nimeoa/olewa kwa miezi 12 iliyopita ; Akendo /okenda ei dweche 12 mokadho)*

☐ Ended a sexual relationship (*Tuliachana na uhusiano wa ngono/ Awera gi jaherana/osiepna)*

☐ Divorced/Separated (*Talaka au kutengana; Ne wawere)*

☐ Widow/widower (*Mjane au mgane/an chi liel/ jaoda ne ose tho)*

**5. Have you been coerced to have sex against your will by your spouse/partner in the last 12 months?** *(Je, umewahi lazimishwa kishiriki ngono kinyume na hiari yako kwa miezi 12 iliyopita/ Bende ose chuni bede e achiel kuom dweche 12 mokalo?)*

☐ Yes (*Ndio/ Kamano)* ☐ No (*La/ Ooyo)*

**6. Did you use a condom consistently with your sexual partner(s) in the last 12 months? *(Tick all that apply)*** *(Jee, ulitumia kondomu kila mara na mshiriki wako wa ngono kwa miezi 12 iliyopita; Bende itiyo gi rabo yunga seche duto ma ibede achiel gi jahera ka ta joherani kuom dweche 12 mosekalo?)*

- ☐ With regular sexual partner(s) ☐ Yes (*Ndio/ Kamano)* ☐ No (*La/Ooyo)*
- ☐ With casual sexual partner(s) ☐ Yes (*Ndio/Kamano)* ☐ No (*La/Ooyo)*
- ☐ Not had sex in the last 12 months *(sijashiriki ngono kwa miezi 12 iliyopita/pok abedo e achiel kuom dweche 12 mokadho)*

**7. Have you had sex under the influence of alcohol and Substance Abuse (ADSA) in the last 12 months?** *(Je, umewahi shiriki ngono kwa ushawishi wa pombe na madawa ya kulevya kwa miezi 12 iliyopita/ Bende isebedo e achiel nikech imer kuom dweche 12 mokalo?*

☐ Yes (*Ndio/ Kamano)* ☐ No (*La/ Ooyo)*

**8. Have you engaged in sex in exchange for money/other favors/in kind (giving or receiving) in the past 12 months?/** *Je, umewahi kushiriki ngono kwa sababu ya malipo au manufaa ya aina yoyote kwa miezi 12 iliyopita?/ Bende ise bedo e achiel nikech chudo kata kony moro amora kuom dweche 12 mosekalo?*

☐ Yes (*Ndio/Kamano)* ☐ No (*La/Ooyo)*

**9. Have you engaged in any of the following in the last 12 months? (*Tick all that apply*)** *(Je, umeshiriki katika yoyote yafuatayo katika miezi 12 iliyopita? ( chagua yeyote umeshiriki); Bende ise timo achiel kuom gigo ma adwa penji kuom dweche 12 mokadho?)*

- ☐ Sex in exchange for favours (*Kushiriki ngono kwa sababu ya manufaa yoyote; Bede achiel nikech yuto moro amora)*
- ☐ Sex work (*Kazi ya ngono/Ohala mar del) (Probe for FSW, MSW, typology- Street based based, bar based, home based, brothel based*
- ☐ Injected drugs for recreation/pleasure (*Kujidunga madawa ya kulevya kwa ajili ya burudani / Chworuok gi yedhe mamero ji)*
- ☐ If male, had sex with men Kama *(kiume, alishiriki ngono na wanaume, Ushoga / Ka dichuo, Bedo achiel gi wuowi)*
- ☐ If female, had anal sex *(Kama kike, amefirwa/ riwruok gi ka siandani)*
- ☐ I have not had sex in the last 12 months *(sijashiriki ngono kwa miezi 12 iliyopita/pok abedo e achiel kuom dweche 12 mokadho)*

**10. Are you currently being treated or have you been treated for STI in the last 12 months? *(Probe for any abnormal urethral or vaginal discharge and genital wounds)*** *(Je, umewahi tibiwa ugonjwa wa zinaa kwa miezi 12 iliyopita ama unatibiwa ugonjwa wa zinaa kwa sasa?/Bende ise yudo thieth mar nyeye kuom dweche 12 mosekalo kata iyudo thieth sani.)*

☐ Yes *(Ndio/Kamano)* ☐ No *(La/Ooyo)*

**11. Marital Status** *(Hali ya ndoa/Bende ise dhi tedo kata ise kendo?)*

- ☐ Never married (*Sijawahi oa au olewa/sijawahi oa/Pok adhi tedo/ Pok akendo)*
- ☐ Married monogamous (*Nimeoa mke mmoja / an gi dhako achiel)*
- ☐ Married polygamous (*Nimeowa wake/a ja doho)*
- ☐ Widow/widower (*Mjane au mgane/An chi liel/ Min oda ne onindo)*
- ☐ Separated/ divorced (*Kutengana/Talaka/ Ne awera gi jaoda)*

**12. Do you live/stay with your spouse/sexual partner?** *(Je, Unaishi na mmeo au mkeo/mshiriki wa ngono/ Bende idak kanyo achiel gi jaodi kata jaherani)*

- ☐ Yes, I live/stay with my spouse/sexual partner (*Ndiyo, nakaa na mpenzi wangu au mshiriki wa ngono/ Kamano)*
- ☐ No, my spouse/sexual partner lives/stays elsewhere or travels away regularly for some days of the week/month/year *(La, mke wangu au mme wangu au mpenzi wangu anaishi mahali pengine/ ooyo, ok wadak kanyakla )*
- ☐ I do not have a partner *(La sijaolewa au sijaoa au sina mpenzi, aonge jaot)*

**13. Are you circumcised (men only)** *(Je, umetahiriwa (waumee pekee)/Ose teri nyangu)*

☐ Yes (*Ndio /Kamano)* ☐ No (*La/Ooyo)* ☐ *Declined to answer*

## Appendix 2: List of Variables Collected as Part of HIV Testing Service Provision

County

Sub-county

Date of visit

Patient type- hospital patient, non-patient

Department- out-patient (OPD), in-patient (IPD)

Type of OPD/IPD visit- new, revisit

Sex- male, female

Age

Tested for HIV before- yes, no

Date of last HIV test

Test report of previous last test- positive, negative

Verified HIV results- yes, no

Eligible for testing- yes, no

Tested for HIV- yes, no

Result- positive, negative

Linked to Care

Reasons for eligibility/ ineligibility

## Appendix 3: List of sub-locations in Siaya County

| | | | | |
|---|---|---|---|---|
| ABOM | KAPIYO | MUNGAO | ONYINYORE | USIGU |
| AJIGO | KARADOLO EAST | MUR_MALANGA | OTHACH | UTONGA |
| AKOM | KARAPUL | MUR_NGIYA | PALA | UYAWI |
| AMBIRA | KATHIENO 'A' | NAYA | PAP ORIANG | UYUNDO |
| ANYIKO | KATHIENO 'B' | NDEDA/OYAMO | RACHAR | WAGAI EAST |
| ANYIKO_YALA | KATHIENO 'C' | NDERE | RAGENG'NI | WAGAI WEST |
| ASAYI | KAUGAGI HAWINGA | NDIGWA | RAMULA | WEST ASANGO |
| BAR CHANDO | KAUGAGI UDENDA | NDORI | RAMUNDE | WEST KARADOLO |
| BAR OLENGO | KOBONG' | NGUGE | RANDOGO | WEST KATWENGA |
| BAR OSIMBO | KOCHIENG 'A' | NGUNYA | RANGALA | WEST MIGWENA |
| BAR SAURI | KOCHIENG 'B' | NORTH RAMBA | RERA | YENGA |
| BAR-AGULU | KODIERE | NORTH RAMBULA | SEGA | YIRO EAST |
| BARDING | KOKWIRI | NYABEDA | SIFUYO EAST | YIRO WEST |
| BAR-KOWINO | KOMENYA KALAKA | NYABERA | SIFUYO WEST | BAR KANYANGO[a] |
| DIENYA EAST | KOMENYA KOWALA | NYADORERA 'A' | SIGER | GONGO[a] |
| DIENYA WEST | KOMOLO | NYADORERA 'B' | SIGOMA URANGA | KAGILO[a] |
| DOHO EAST | KOYEYO | NYAGOKO | SIGOMRE | KANYADET[a] |
| DOHO WEST | KUKUMU_KOMBEWA | NYAJUOK | SIMENYA | KAUDHA EAST[a] |
| EAST ASANGO | LIETA | NYALENYA | SIMUR | KAUDHA WEST[a] |
| EAST KATWENGA | LIGALA NORTH | NYALGUNGA | SIMUR EAST | MADUNGU[a] |
| EAST MIGWENA | LIGEGA | NYAMILA | SIMUR KONDIEK | MAHANGA[a] |
| GANGU | LIHANDA | NYAMNINIA | SIRANGA | MAHOLA_ULAWE[a] |
| GOMBE | LUNDHA | NYAMSENDA | SIREMBE | MALUNGA CENTRAL[a] |
| GOT ABIERO | MAGOYA | NYANDIWA | SIRIWO | MITUNDU[a] |
| GOT AGULU | MAHAYA | NYANDIWA_YALA | SOUTH RAMBA | NDENGA[a] |
| GOT OSIMBO | MALANGA | NYANGOMA | SOUTH RAMBULA | NYAGUDA[a] |
| GOT RAMOGI | MALIERA | NYANGOMA_ALEGO | SUMBA | NYAMONYE[a] |
| GOT REGEA | MALUNGA EAST | NYAWARA | TINGARE EAST | RUWE[a] |
| HONO | MALUNGA WEST | NYAWITA | TINGARE WEST | SIHAY[a] |
| JERA | MARANDA | OBAMBO | UGUNJA | UHUYI[a] |
| JINA | MARENYO | OCHIENG'A | ULAFU | USIRE[a] |
| KABURA UHUYI | MASALA | OJWADO 'B' | ULAMBA | |
| KAGONYA | MASAT EAST | OJWANDO 'A' | UMALA | |
| KAGWA | MASAT WEST | OLWA | UMALA_UGUNJA | |
| KALKADA URADI | MASUMBI | OMIA DIERE | URANGA | |
| KAMBARE | MEMBA | OMIA MALO | URIRI | |
| KANDENGE | MULAHA | OMIA MWALO | USENGE | |
| [a]Sub-locations not included in the analysis | | | | |

**Appendix 4: Variables Collected in the Home-based HIV Testing Enumeration Form**

County

Sub-county

Ward

Village

Homestead/Compound

Relationship to head- homestead head, spouse, children, parent, in-law, co-wife, sibling, other relative

Sex- male, female

Age

**Appendix 5: Variables Collected in Ministry of Health HIV testing registers during Home-based HIV Testing**

Date

Age

Sex- male, female

Tested before

Previous HIV test results

When last tested- (<12 months; >12 months)

Marital status- single, married monogamous, married polygamous, separated/divorced, widow/widower

Key population- general population, female sex worker, fisher folk

Eligible for HIV testing

Tested for HIV

Final HIV results

Discordant couple

Linked to HIV clinic

## Appendix 6: Maseno University Ethics Review Committee Approval Letter

# MASENO UNIVERSITY ETHICS REVIEW COMMITTEE

Tel: +254 057 351 622 Ext: 3050
Fax: +254 057 351 221
Private Bag – 40105, Maseno, Kenya
Email: muerc-secretariate@maseno.ac.ke

**FROM**: Secretary - MUERC

**DATE**: 4h July, 2019

**TO**: Dr. Kawango Agot and Dr. Ohaga Spala
Impact Research Development and Organization
Baring Road, Milimani Estate Kisumu
P.O. Box 9171-40141, Kisumu, Kenya

**REF**:MSU/DRPI/MUERC/00700/19

**RE: Utilization of Routinely Collected HIV Prevention, Care and Treatment Data to Improve Program Implementation in Siaya, Homabay and Busia Counties, Kenya. Proposal Reference Number MSU/DRPI/MUERC/00700/19**

This is to inform you that the Maseno University Ethics Review Committee (MUERC) determined that the ethics issues raised at the initial review were adequately addressed in the revised proposal. Consequently, the study is granted approval for implementation effective this 4th day of July, 2019 for a period of one (1) year. This is subject to getting approvals from NACOSTI and other relevant authorities.

Please note that authorization to conduct this study will automatically expire on 3rd July, 2020. If you plan to continue with the study beyond this date, please submit an application for continuation approval to the MUERC Secretariat by 15th June, 2020.

Approval for continuation of the study will be subject to successful submission of an annual progress report that is to reach the MUERC Secretariat by 15th June, 2020.

Please note that any unanticipated problems resulting from the conduct of this study must be reported to MUERC. You are required to submit any proposed changes to this study to MUERC for review and approval prior to initiation. Please advice MUERC when the study is completed or discontinued.

Thank you.

MASENO UNIVERSITY SECRETARY 04 JUL 2019 ETHICS REVIEW COMMITTEE

Dr. Bernard Guyah
Ag. Secretary,
Maseno University Ethics Review Committee.

Cc: Chairman,
Maseno University Ethics Review Committee.

## Appendix 7: Kenyatta National Hospital- University of Nairobi Ethics Approval Letter.

UNIVERSITY OF NAIROBI
COLLEGE OF HEALTH SCIENCES
P O BOX 19676 Code 00202
Telegrams: varsity
Tel:(254-020) 2726300 Ext 44355

KNH-UON ERC
Email: uonknh_erc@uonbi.ac.ke
Website: http://www.erc.uonbi.ac.ke
Facebook: https://www.facebook.com/uonknh.erc
Twitter: @UONKNH_ERC https://twitter.com/UONKNH_ERC

KENYATTA NATIONAL HOSPITAL
P O BOX 20723 Code 00202
Tel: 726300-9
Fax: 725272
Telegrams: MEDSUP, Nairobi

Ref: KNH-ERC/A/432

6th December 2018

KENYATTA NATIONAL HOSPITAL
APPROVED
06 DEC 2018
KNH/UoN-ERC
P. O. BOX 20723 - 00202 NAIROBI

Wanjohi Stella
Principal Investigator
HIV Prevention and Community Services
Centre for Health Solutions, Kenyan
P O BOX 62729-00200
NAIROBI
Email: swanjohi@chskenya.org

Dear Stella

**RESEARCH PROPOSAL – ASSESSMENT OF EXPANDED HIV TESTING ELIGIBILITY AT HIGH-VOLUME PEPFAR-SUPPORTED FACILITIES IN WESTERN KENYA (P513/07/2018)**

This is to inform you that the KNH- UoN Ethics & Research Committee (KNH- UoN ERC) has reviewed and **approved** your above research proposal. The approval period is 6th December 2018 – 5th December 2019. The waiver of informed consent is hereto approved.

This approval is subject to compliance with the following requirements:

a) Only approved documents (informed consents, study instruments, advertising materials etc) will be used.
b) All changes (amendments, deviations, violations etc.) are submitted for review and approval by KNH-UoN ERC before implementation.
c) Death and life threatening problems and serious adverse events (SAEs) or unexpected adverse events whether related or unrelated to the study must be reported to the KNH-UoN ERC within 72 hours of notification.
d) Any changes, anticipated or otherwise that may increase the risks or affect safety or welfare of study participants and others or affect the integrity of the research must be reported to KNH- UoN ERC within 72 hours.
e) Clearance for export of biological specimens must be obtained from KNH- UoN ERC for each batch of shipment.
f) Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. (*Attach a comprehensive progress report to support the renewal*).
g) Submission of an *executive summary* report within 90 days upon completion of the study.
This information will form part of the data base that will be consulted in future when processing related research studies so as to minimize chances of study duplication and/ or plagiarism.

For more details consult the KNH- UoN ERC website **http://www.erc.uonbi.ac.ke**

Yours sincerely,

**PROF. M.L. CHINDIA**
**SECRETARY, KNH-UoN ERC**

c.c. The Principal, College of Health Sciences, UoN
The Director, CS, KNH
The Chairperson, KNH-UoN ERC
The Assistant Director, Health Information, KNH
Co-investigators: Rachel Joseph (CDC Kenya) Kevin DeCock(CDC-Kenya), Lucy Ng'ang'a(CDC, Kenya), Hellen Muttai (CDC,Kenya), Paul Musingila (CDC, Kenya),Fredrick Miruka(CC Kenya), Gordon Okomo (Homa Bay County Director of Health), Iscah Moth(Homa Bay County), Dickens Onyango(Kisumu County Director of Health),Eunice Kinywa (Kisumu County),Samuel Omondi(Siaya County Director of Health),Caren Ayieko(Siaya County),Fillet Lugali(ICAP),Caroline O.Dande(KEMRI-UCSF,Kisumu),Polycarp Musee(EGPAF), Emily Zielinski-Gutierrez(CDC)

# Appendix 8: Centers for Disease Control and Prevention Ethics Approval Letter

CGH HSR Tracking #:_____

## Request for Project Determination & Approval -- Center for Global Health (CGH)

Use this form to submit proposals to the CGH Office of the Associate Director for Science/Laboratory Science (ADS/ADLS) for research/nonresearch determination and requirements for IRB review/approval.
**Approval Chain: Investigator → Branch Chief/Country Director →Division ADS → CGH Human Subjects Mailbox**

☑ **New Request** ☐ **Amendment** ☐ **Laboratory Submission**

| | | |
|---|---|---|
| Project Title: Utilization of Routinely Collected HIV Prevention Care and Treatment Data to Improve Program Implementation in Siaya, Homa Bay and Busia Counties, Kenya [#D.19.2018] | | Project Location/Country(ies): Siaya, Busia and Homa Bay Counties of Kenya |
| CDC Principal Investigator's name and SEV#: Kawango Agot. PhD, MPH | CDC Primary Contact's name and SEV# (Leave blank if same as PI): Elijah Odoyo-June. PhD [SEV# 19947] | |
| Division: DGHT | CDC PI or PC Email: yed0@cdc.gov | Telephone: +254 724 256805 |
| Project start date (mm/dd/yyyy): 09/30/2016 | Project end date (mm/dd/yyyy): 09/29/2023 | |

**Collaborating Institutions** (List other collaborating institutions in the protocol or in a separate document)

| | | |
|---|---|---|
| ☑CoAg ☐Grant ☐contract #: 1NU2GGH001963 | Original Award Year if CoAg: 2016 | Current Budget Year if CoAg: 2018 |
| Title (CoAg, Grant, or Contract): Increasing Access to and Availability of Sustainable, High Quality, Comprehensive Health and Structural Interventions a | | |
| Supported Institution Name: Impact Research and Development Organization [IRDO] | | |
| Supported Institution FWA# (if applicable): 00014906 | FWA Exp. Date (if applicable): 05/23/2022 | |

**Check appropriate category and subcategory**

☐ **I. Activity is NOT human subjects research. Primary intent is public health practice or a disease control activity (Check all that apply)**
- ☐ **A.** Epidemic or endemic disease control activity; if applicable, Epi-AID #
- ☐ **B.** Routine surveillance activity (e.g., disease, adverse events, injuries)
- ☐ **C.** Program evaluation activity*
- ☐ **D.** Public health program activityΩ
- ☐ **E.** Laboratory proficiency testing

* Evaluation of a new intervention for effectiveness and comparison of different interventions are research under CDC policy.
Ω e.g., service delivery; health education programs; social marketing campaigns; program monitoring; electronic database construction and/or support; development of patient registries; needs assessments; and demonstration projects intended to assess organizational needs, management, and human resource requirements for implementation.

☐ **II. Activity is research but does NOT involve human subjects (Check all that apply)**
- ☐ **A.** Activity is research involving collection or analysis of data about health facilities or other organizations or units (NOT persons).
- ☐ **B.** Activity is research involving data or specimens from deceased persons.
- ☐ **C.** Activity is research involving unlinked or anonymous data or specimens collected for another purpose.
- ☐ **D.** Activity is research involving data or specimens from animal subjects. §

§Note: Approval by CDC Institutional Animal Care and Use Committee (IACUC) may be required for certain animal research. Institution must also have assurance with the Office of Laboratory and Animal Welfare at NIH.

☑ **III. Activity is research involving human subjects but CDC involvement does not constitute "engagement in human subject research."** CDC employees or agents will not intervene or interact with living individuals or have access to identifiable information for research purposes. Appropriate IRB or ethics committee approval is required prior to approval. **(Check all that apply)**
- ☑ **A.** This project is funded under a grant/cooperative agreement/contract award mechanism.
- ☐ **B.** CDC staff provide technical support that does not involve possession or analysis of identifiable data or interaction with participants from whom data are being collected (No CDC Supportβ).
- ☐ **C.** CDC staff are involved only in manuscript writing for a project that has closed. For the project, CDC staff did not interact with participants and were not involved with data collection (No CDC Support).
- ☐ **D.** Activity is research involving linked data, but CDC non-disclosure form 0.1375B is signed.∞

β See definition of support on page 3.
∞ CDC form 0.1375B agreement is required for all subcategories (A-D) if CDC has access to linked data. This agreement prohibits the release of identifying key to CDC investigators under any circumstances. The purposes of the planned research do not contradict the terms of consent under which the information or specimens were collected, whether that consent was documented or not documented.

☐ **IV. Activity is research involving human subjects that requires submission to CDC Human Research Protection Office (Check one)α**
- ☐ A. Full Board Review (Use forms 0.1250, 0.1370-research partners)
- ☐ B. Expedited Review (Use same forms as A above)
- ☐ C. Exemption Request¥ (Use forms 0.1250X, 0.1370-research partners)
- ☐ D. Reliance¥
  - ☐ 1. Request to allow CDC to rely on a non-CDC IRB (Use same forms as A above, **plus** 0.1371)
  - ☐ 2. Request to allow outside institution to rely on CDC IRB (Use same forms as A above, plus 0.1372)

α There are other types of requests not listed under category IV, e.g., continuation of existing protocol, amendment, incident reports.
¥ Exemption and reliance request is approved by CDC Human Research Protection Office (HRPO).

CGH HSR Tracking #:_____

## Public Access and Data Sharing

**A. Type of data collected or generated:** 2. By non-CDC staff, supported with CDC funding

***Instructions:*** From the dropdown list, select the types of data that will be collected that best fits this project. Categories 1, 2, and 3 are data covered by CDC Policy (http://aops-mas-iis/Policy/Doc/policy385.pdf). Categories 4 and 5 are data covered by CDC Policy but release or sharing may be restricted or limited. Categories 6, 7 and 8 are data NOT covered by CDC Policy and no further information is needed under this section. Use the lowest number when the data falls under more than one type. **See Box below for more information on the categories.**

*Provide a 2-3 sentences description of the data that will be collected in this project:*

Through CDC funding, IRDO is implementing HIV prevention programs including evidence-based package of HIV prevention, care and treatment interventions to HIV vulnerable populations in Siaya, Homabay and Busia Counties of Kenya. Routine program data collected during service delivery will be analyzed to identify

**B. Data ownership:** Name of Organization: Impact Research and Development Organization

***Instructions:*** Provide the name of the organization that will own the public health data for this project. If there are multiple organizations involved, provide the name of the organization that will retain and provide long-term control over the access and use of the data. Provide data steward's name and contact information if available.

**C. Public access level:** 1. Public – dataset or summary data tables will be available without restriction after data availability date

***Instructions:*** From the drop-down list, select the data release category that will best fits how data will be available after data availability date.

**Justifications:**

***Instructions:*** From the dropdown list, select the option that best fits the justification for restricted access or unavailability of data.

*Provide a brief description (1-3 sentences) if "Other reason" is selected:*

N/A

**D. Anticipated data availability date (if applicable):** 10/2021

***Instructions:*** Provide the anticipated date (mm/yyyy) that the public health data will become available.

CGH HSR Tracking #:_____

**Amendment**

If this request is an amendment to an existing project determination. Please include a brief description of the substantive change or modification below and attach both clean and marked copies of the amended protocol or project outline.

N/A

**Submission:** Attach a protocol if one exists. If not, provide a separate project description (See suggested format below) in sufficient detail to justify the proposed category. Submit your request to your branch chief (or country director or designee for country staff).

| **CGH ADS/ADLS Review** | **Date received in CGH ADS /ADLS office:** |
|---|---|

☐ **Project does not require human subject research review beyond CGH at this time. Local IRB Exp. Date (if applieable):**

☐ **Project constitutes human subject research that must be routed to CDC HRPO.**

Comments/Rationale for Determination:

**Approvals and Signatures**

| | **Date:** | **Remarks:** |
|---|---|---|
| Dr. Kawango Agot Digitally signed by Dr. Kawango Agot Date: 2018.08.08 16:57:28 +03'00' Investigator | 08/08/2018 | |
| Cathy Toroitich-Ruto, PhD, CCRA Digitally signed by Cathy Toroitich-Ruto, PhD, CCRA Date: 2018.08.23 16:49:37 +03'00' Branch Chief/Country Director | | |
| Division Human Research Protection Coordinator Division ADS/ADLS or Director | 5/9/19 | Pend, local IRB approval, |
| CGH Human Research Protection Coordinator CGH ADS/ADLS or Deputy ADS/ADLS | | |

Note: Although CDC IRB review is not required for certain projects (categories I,II & III) approved under this determination, CDC investigators and project officers are expected to adhere to the highest ethical standards of conduct and to respect and protect to the extent possible the privacy, confidentiality, and autonomy of participants. All applicable country, state, and federal laws must be followed. Informed consent may be appropriate and should address all applicable elements of informed consent. CDC investigators should incorporate diverse perspectives that respect the values, beliefs, and cultures of the people in the country, state, and community in which they work.

# Appendix 9: Centers for Disease Control and Prevention Ethics Approval Letter

CGH HSR Tracking #:_____

## Request for Project Determination & Approval -- Center for Global Health (CGH)

Use this form to submit proposals to the CGH Office of the Associate Director for Science/Laboratory Science (ADS/ADLS) for research/nonresearch determination and requirements for IRB review/approval.
**Approval Chain: Investigator → Branch Chief/Country Director →Division ADS → CGH Human Subjects Mailbox**

☑ New Request ☐ Amendment ☐ Laboratory Submission

| | |
|---|---|
| Project Title: Assessment of Expanded HIV Testing Eligibility at High-volume PEPFAR-supported Facilities in western Kenya (D-12-2018) | Project Location/Country(ies): Kenya |
| CDC Principal Investigator's name and SEV#: | CDC Primary Contact's name and SEV# (Leave blank if same as PI): Rachael Joseph SEV#7220 |
| Division: DGHT | CDC PI or PC Email: vle5@cdc.gov / Telephone: +254 722 209 715 |
| Project start date (mm/dd/yyyy): 03/01/2017 | Project end date (mm/dd/yyyy): 03/01/2020 |

**Collaborating Institutions** (List other collaborating institutions in the protocol or in a separate document)

| | |
|---|---|
| ☑ CoAg ☐ Grant ☐ contract #: NU2GGH001946-01-00 / Original Award Year if CoAg: 2016 | Current Budget Year if CoAg: YR 3 |
| Title (CoAg, Grant, or Contract): Shinda Project: Implementation and Expansion of High Quality HIV Care, Prevention and Treatment | |
| Supported Institution Name: Centre for Health Solutions, Kenya (CHS) | |
| Supported Institution FWA# (if applicable): FWA00021277 | FWA Exp. Date (if applicable): 01/23/2019 |

**Check appropriate category and subcategory**

☐ **I. Activity is NOT human subjects research. Primary intent is public health practice or a disease control activity (Check all that apply)**
- ☐ A. Epidemic or endemic disease control activity; if applicable, Epi-AID #
- ☐ B. Routine surveillance activity (e.g., disease, adverse events, injuries)
- ☐ C. Program evaluation activity*
- ☐ D. Public health program activityΩ
- ☐ E. Laboratory proficiency testing

* Evaluation of a new intervention for effectiveness and comparison of different interventions are research under CDC policy.
Ω e.g., service delivery; health education programs; social marketing campaigns; program monitoring; electronic database construction and/or support; development of patient registries; needs assessments; and demonstration projects intended to assess organizational needs, management, and human resource requirements for implementation.

☐ **II. Activity is research but does NOT involve human subjects (Check all that apply)**
- ☐ A. Activity is research involving collection or analysis of data about health facilities or other organizations or units (NOT persons).
- ☐ B. Activity is research involving data or specimens from deceased persons.
- ☐ C. Activity is research involving unlinked or anonymous data or specimens collected for another purpose.
- ☐ D. Activity is research involving data or specimens from animal subjects. §

§Note: Approval by CDC Institutional Animal Care and Use Committee (IACUC) may be required for certain animal research. Institution must also have assurance with the Office of Laboratory and Animal Welfare at NIH.

☑ **III. Activity is research involving human subjects but CDC involvement does not constitute "engagement in human subject research."** CDC employees or agents will not intervene or interact with living individuals or have access to identifiable information for research purposes. Appropriate IRB or ethics committee approval is required prior to approval. **(Check all that apply)**
- ☑ A. This project is funded under a grant/cooperative agreement/contract award mechanism.
- ☑ B. CDC staff provide technical support that does not involve possession or analysis of identifiable data or interaction with participants from whom data are being collected (No CDC Supportβ).
- ☐ C. CDC staff are involved only in manuscript writing for a project that has closed. For the project, CDC staff did not interact with participants and were not involved with data collection (No CDC Support).
- ☐ D. Activity is research involving linked data, but CDC non-disclosure form 0.1375B is signed.ω

CM 11/1/18

β See definition of support on page 3.
ω CDC form 0.1375B agreement is required for all subcategories (A-D) if CDC has access to linked data. This agreement prohibits the release of identifying key to CDC investigators under any circumstances. The purposes of the planned research do not contradict the terms of consent under which the information or specimens were collected, whether that consent was documented or not documented.

☐ **IV. Activity is research involving human subjects that requires submission to CDC Human Research Protection Office (Check one)α**
- ☐ A. Full Board Review (Use forms 0.1250, 0.1370-research partners)
- ☐ B. Expedited Review (Use same forms as A above)
- ☐ C. Exemption Request¥ (Use forms 0.1250X, 0.1370-research partners)
- ☐ D. Reliance¥
  - ☐ 1. Request to allow CDC to rely on a non-CDC IRB (Use same forms as A above, plus 0.1371)
  - ☐ 2. Request to allow outside institution to rely on CDC IRB (Use same forms as A above, plus 0.1372)

α There are other types of requests not listed under category IV, e.g., continuation of existing protocol, amendment, incident reports.
¥ Exemption and reliance request is approved by CDC Human Research Protection Office (HRPO).

CGH HSR Tracking #:_____

## Public Access and Data Sharing

**A. Type of data collected or generated:** 2. By non-CDC staff, supported with CDC funding

***Instructions:*** From the dropdown list, select the types of data that will be collected that best fits this project. Categories 1, 2, and 3 are data covered by CDC Policy (http://aops-mas-nt/Policy/Doc/policy385.pdf). Categories 4 and 5 are data covered by CDC Policy but release or sharing may be restricted or limited. Categories 6, 7 and 8 are data NOT covered by CDC Policy and no further information is needed under this section. Use the lowest number when the data falls under more than one type. **See Box below for more information on the categories.**

*Provide a 2-3 sentences description of the data that will be collected in this project:*

De-identified routinely collected HIV testing data from standardized Kenya Ministry of Health registers; behavioral risk factor information collected during routine HIV testing and counseling offered in facility outpatient departments.

**B. Data ownership:** Kenya Ministry of Health

***Instructions:*** Provide the name of the organization that will own the public health data for this project. If there are multiple organizations involved, provide the name of the organization that will retain and provide long-term control over the access and use of the data. Provide data steward's name and contact information if available.

**C. Public access level:** 2. Restricted public – dataset will be available to the public under certain use restrictions after data availability date

***Instructions:*** From the drop-down list, select the data release category that will best fits how data will be available after data availability date.

**Justifications:** 1. Country owns the data with protections under their laws and regulations

***Instructions:*** From the dropdown list, select the option that best fits the justification for restricted access or unavailability of data.

*Provide a brief description (1-3 sentences) if "Other reason" is selected:*

**D. Anticipated data availability date (if applicable):** 03/2019

***Instructions:*** Provide the anticipated date (mm/yyyy) that the public health data will become available

### Box: Type of data collected or generated

1. *By CDC staff, supported with CDC funding* – CDC funds the activity and data are collected by CDC staff.
2. *By non-CDC staff, supported with CDC funding* – CDC funds or co-funds the data collection through mechanisms such as grants, cooperative agreements, contracts, or other funding mechanisms. Data is collected by non-CDC staff. When CDC funds another federal agency, an interagency agreement should indicate who would be responsible for the data.
3. *Provided to CDC and becomes part of a CDC data system* – Data is reported to CDC by another entity, e.g., by local health departments, that become a part of a CDC data collection system, e.g., CDC surveillance systems.
4. *Owned by partner and protected from release by laws or regulations* – CDC funds or co-funds the data collection. Data may be collected by CDC or non-CDC staff as in #1 and 2 above, but applicable US or country laws and regulations limit or restrict disclosure of data. Examples of US laws limiting disclosure include: the Privacy Act, Trade Secrets Act, or Section 308(d) of the US Public Health Service Act. When CDC funds activities in other countries, foreign laws and/or regulations may also apply.
5. *Not sharable due to potential dual-use research of concern* – Dual use research of concern is life sciences research that, based on current understanding, can reasonably be anticipated to provide knowledge, products, or technologies that could be directly misapplied by others to pose a threat to public health, agriculture, plants, animals, the environment, or material (http://intranet.cdc.gov/padiss/manuals-and-policies/dual-use-research/). Data pertaining to DUR may not be sharable because of the potential threats.
6. *Owned by partner and shared with CDC, but no CDC funding* – CDC does not fund, collect, or own the data. Data may be shared with CDC.
7. *Owned by partner and has an agreement restricting data sharing* – CDC does not fund, collect, or own the data. Data may be shared with CDC under data sharing agreement.
8. *By another federal agency* – Another federal agency shares data with CDC under restricted terms agreement. The other federal agency is responsible for data release and sharing.

CGH HSR Tracking #:_____

**Amendment**

If this request is an amendment to an existing project determination. Please include a brief description of the substantive change or modification below and attach both clean and marked copies of the amended protocol or project outline.

**Submission:** Attach a protocol if one exists. If not, provide a separate project description (See suggested format below) in sufficient detail to justify the proposed category. Submit your request to your branch chief (or country director or designee for country staff).

| CGH ADS/ADLS Review | Date received in CGH ADS /ADLS office: |
|---|---|

☐ **Project does not require human subject research review beyond CGH at this time. Local IRB Exp. Date (if applicable):**

☐ **Project constitutes human subject research that must be routed to CDC HRPO.**

Comments/Rationale for Determination:

**Approvals and Signatures**

| | Date: | Remarks: |
|---|---|---|
| Investigator | 10/11/18 | |
| CATHY TORRITICH-RUTO<br>Branch Chief/Country Director | 10/12/2018 | |
| Division Human Research Protection Coordinator<br>Division ADS/ADLS or Director | 11/1/18 | pending local IRB approval. |
| CGH Human Research Protection Coordinator<br>CGH ADS/ADLS or Deputy ADS/ADLS | | |

Note: Although CDC IRB review is not required for certain projects (categories I,II & III) approved under this determination, CDC investigators and project officers are expected to adhere to the highest ethical standards of conduct and to respect and protect to the extent possible the privacy, confidentiality, and autonomy of participants. All applicable country, state, and federal laws must be followed. Informed consent may be appropriate and should address all applicable elements of informed consent. CDC investigators should incorporate diverse perspectives that respect the values, beliefs, and cultures of the people in the country, state, and community in which they work.