ECONOMIC EVALUATION OF NATIONALLY SCALED POINT-OF-CARE DIAGNOSTIC PLATFORMS FOR VIRAL LOAD MONITORING IN KENYA

September 2018
This publication was produced for review by the United States Agency for International Development.
It was prepared by Emily Mangone, Christopher Cintron, Ben Johns, Victor Mudhune, and Carlos Avila for the Health Finance and Governance Project

The Health Finance and Governance Project
USAID's Health Finance and Governance (HFG) project will help to improve health in developing countries by expanding people's access to health care. Led by Abt Associates, the project team will work with partner countries to increase their domestic resources for health, manage those precious resources more effectively, and make wise purchasing decisions. As a result, this five-year, $209 million global project will increase the use of both primary and priority health services, including HIV/AIDS, tuberculosis, malaria, and reproductive health services. Designed to fundamentally strengthen health systems, HFG will support countries as they navigate the economic transitions needed to achieve universal health care.

September 2018

Cooperative Agreement No: AID-OAA-A-12-00080

Submitted to: Scott Stewart, AOR
Office of Health Systems
Bureau for Global Health

Recommended Citation:
Mangone ER, Cintron C, Johns B, Mudhune V, Avila C. 2018. *Economic Evaluation of Nationally Scaled Point-of-Care Diagnostic Platforms for Viral Load Monitoring in Kenya*. Rockville, MD: Health Finance and Governance Project, Abt Associates Inc.

Abt Associates Inc. | 6130 Executive Boulevard | Rockville, Maryland 20852
T: 301.347.5000 | F: 301.652.3916 | www.abtassociates.com

Broad Branch Associates | Development Alternatives Inc. (DAI) | Futures Institute | Johns Hopkins Bloomberg School of Public Health (JHSPH) | Results for Development Institute (R4D) | RTI International | Training Resources Group, Inc. (TRG)

ECONOMIC EVALUATION OF NATIONALLY SCALED POINT-OF-CARE DIAGNOSTIC PLATFORMS FOR VIRAL LOAD MONITORING IN KENYA

CONTENTS

Contents ..... **i**

Acronyms **3**

Executive Summary **4**

1. Introduction **6**

1.1 Background and Context 6
1.2 Objective 11

2. Methodology **12**

2.1 Model Overview 12
2.2 Transition Parameters 17
2.3 Costs 21
2.4 Disability Weights 23
2.5 Discounting 24
2.6 Sensitivity Analysis 24

3. Results **26**

3.1 Model Outcomes 26
3.2 Deterministic Sensitivity Analysis 28
3.3 Probabilistic Uncertainty Analysis 29

4. Discussion **33**

4.1 Key Takeaways 33
4.2 Limitations 34
4.3 Conclusion 35

References **36**

List of Tables

Table 1: Transition Parameters 17
Table 2: Component Costs for Laboratory and POC VL Testing 22
Table 3: Disability Weights for HIV Health States (GBD 2016) 23
Table 4: Disability Weights for Model Health States 24
Table 5: Sensitivity Analysis Parameter Values and Distributions 25
Table 6: Discounted Outcomes over 10-Year Horizon 26
Table 7: Incremental Model Outcomes 26
Table 8: Discounted Outcomes over 10-Year Horizon with Added Cohort Entries 27

List of Figures

Figure 1: Central Laboratory Equipment (The Global Fund 2017) 8
Figure 2: Point-of-Care Equipment (The Global Fund 2017) 10
Figure 3: Markov Model Diagram 14
Figure 4. Average Turnaround Timeline from Primary Data Collection 19
Figure 5: POC Incremental Cost-Effectiveness Ratios over 10 years 27
Figure 6: Tornado Diagram for POC Cost per DALY Averted at 10 years 28
Figure 7: Tornado Diagram for POC Cost per HIV Transmission Averted 29
Figure 8: ICER Plane for POC DALYs Averted 30
Figure 9: POC Cost-Effectiveness Acceptability Curve for DALYs Averted 31
Figure 10: Density Graph of ICERs for Transmissions Averted in POC 32

ACRONYMS

AIDS	Acquired immune deficiency syndrome
ART	Antiretroviral therapy
DALY	Disability-adjusted life year
DBS	Dried blood spot
EAC	Enhanced adherence counseling
FLART	First line antiretroviral therapy
GBD	Global burden of disease
GDP	Gross domestic product
HFG	Health Finance and Governance project
HIV	Human immunodeficiency virus
ICER	Incremental cost-effectiveness ratio
KEMRI	Kenya Medical Research Institute
LMIC	Low- and middle-income country
LTFU	Loss/lost to follow-up
MSF	Médecins Sans Frontières
NASCOP	National AIDS and STI Control Programme
POC	Point-of-care
SLART	Second line antiretroviral therapy
TAT	Turnaround time
UNAIDS	The Joint United Nations Programme on HIV/AIDS
USAID	United States Agency for International Development
VL	Viral load
WHO	World Health Organization
WTP	Willingness-to-pay

EXECUTIVE SUMMARY

Introduction

The National AIDS and STI Control Programme (NASCOP) in Kenya has scaled up routine viral load (VL) monitoring to serve the more than one million HIV-positive Kenyans receiving antiretroviral therapy (ART). Yet the current model of batching blood samples to central laboratories presents logistical challenges including delays and loss of test results, and there is limited information on how the cost of central laboratory VL monitoring compares to alternative approaches. Point-of-care (POC) VL testing is a novel approach to monitoring ART adherence and failure which processes individual tests on site with rapid turnaround time between blood draw and results counseling (<120 minutes). POC tests are practical in low-resource settings because they can function without electricity, running water, transport, maintenance, or specialized technicians. In 2017, the World Health Organization (WHO) approved the first quantitative HIV POC VL assay for use in low-resource settings. Given the potential of POC VL monitoring to improve the utility and expediency of VL monitoring, a cost-effectiveness modeling study was conducted to compare the costs and health outcomes of POC and central laboratory VL monitoring approaches in Kenya to inform future investment.

Methodology

A Markov model was developed from a health system perspective to compare routine POC VL monitoring to central laboratory VL monitoring for a hypothetical cohort of 1,000,000 adult (18+) HIV-positive Kenyans on ART. The model follows the cohort over a 10-year time horizon through first and second line treatments using a six-month cycle length to accommodate NASCOP/WHO testing guidelines. Patients were modeled to be at risk of death, loss to follow-up, and increased probability of transmission due to uncontrolled VL. Disability-adjusted life years (DALYs) were calculated over a lifetime analytical time horizon to capture, compare, and monetize health losses. Costs in 2016 U.S. dollars were pulled from a parallel Health Finance and Governance project costing study that evaluated the cost of Cepheid GeneXpert-IV and Alere POC diagnostic platforms and Abbott and Roche laboratory technology for VL monitoring in Kenya and these were supplemented from the scientific literature. Transition parameters were derived from primary data collection on turnaround times and published literature. Costs and outcomes were discounted by 3 percent annually. Univariate sensitivity analyses and probabilistic uncertainty analysis explored the robustness of the results.

Results

POC VL monitoring is cost effective compared to the current approach of central laboratory VL monitoring over a time horizon of 10 years. While POC VL was the more expensive approach to implement, costing USD $37.4 million more than the central laboratory approach over 10 years, it also resulted in fewer DALYs (-13,606), secondary HIV transmissions (-690), and deaths (-855) over 10 years. At a cost of $2,752 per DALY averted, POC VL monitoring was well within the WHO-recommended willingness-to-pay (WTP) threshold of three times the 2016 per capita gross domestic product (GDP) ($4,365). The probabilistic sensitivity analysis supported the base case findings and indicated that the probability that POC was cost effective compared to central laboratory was very high, even given a range of distributional uncertainty. The indifference point of the POC vs. central laboratory approach to VL monitoring in probabilistic uncertainty analysis was a WTP threshold of $1,250, which is

less than the 2016 per capita GDP in Kenya ($1,455). In sensitivity analyses, the model was most sensitive to cost differences in approaches, followed by uncertainty around the probability of delay of test receipts for central laboratory VL monitoring.

Discussion

These findings highlight the fact that while the current central laboratory VL monitoring program functions well, innovative technologies and approaches such as POC VL monitoring can improve the expediency and effectiveness of VL monitoring, resulting in better patient care, clinical outcomes, and HIV control. The Kenyan Ministry of Health and NASCOP leadership should consider piloting a POC VL monitoring approach, particularly in hard-to-reach settings and populations where transport and continuity of care do pose significant obstacles to effective VL monitoring. Pilot data could also be used to make stronger and more precise predictions and recommendations for investments in POC VL monitoring at scale.

1. INTRODUCTION

1.1 Background and Context

1.1.1 Overview

With an HIV prevalence of 5.4 percent among adults 15–49, Kenya has the eleventh highest HIV prevalence globally (UNAIDS 2016). In response to this generalized epidemic, HIV viral load (VL) testing and counseling has become a cornerstone of the Kenyan National AIDS and STI Control Programme's (NASCOP) approach to monitoring antiretroviral therapy (ART) treatment and preventing secondary HIV transmission (NASCOP 2014). Among patients already diagnosed with HIV, the benefits of ART are limited by imperfect adherence and treatment failure, which leads to increases of HIV in the blood and acquired drug resistance (Vandormael et al. 2016). Routine VL testing in this population can support adherence, ensure appropriate treatment, and confirm viral suppression, thereby reducing secondary transmission. However, there is limited research on the most clinically effective and cost-effective approach to VL monitoring, especially given limited financial resources and the large populations receiving ART (over one million adults in Kenya) (Estill et al. 2013, Barnabas et al. 2017).

1.1.2 Viral load monitoring

VL tests determine the amount of HIV particles in a patient's blood, and achievement of an "undetectable" VL indicates that ART is effectively preventing viral replication (WHO 2013). Patients with an elevated VL[1] have a higher risk of drug resistance and treatment failure (WHO 2013, MSF 2014a). Since 2013, VL testing has replaced CD4 count as the recommended gold standard for monitoring and detection of treatment failure among HIV-positive individuals receiving ART (WHO 2013). While there is not yet established evidence of survival benefit of VL testing over CD4 count, evidence suggests that VL testing can offer a more accurate and earlier indication of treatment failure, reduce the amount of time spent failing ART, discern treatment failure from non-adherence, and can be used as a proxy for the risk of population-level transmission (WHO 2013, Salazar-Vizcaya et al. 2014). Further, routine VL testing, as opposed to targeted VL testing, has been shown to significantly reduce the risk of virological failure (Sawe et al. 2013, Kerschberger et al. 2015). Because of these findings, VL testing is a key component of The Joint United Nations Programme on HIV/Acquired Immune Deficiency Syndrome (UNAIDS) "90-90-90" goals, which are to increase to 90 percent by 2020 the proportions of 1) persons living with HIV infection who know their status, 2) persons with diagnosed HIV infection receiving ART, and 3) persons on ART who have achieved viral suppression (UNAIDS 2014a). Consequently, 39 of 52 low- and middle-income countries (LMICs) have developed guidelines for implementing routine VL testing as the preferred monitoring technology for people on ART (MSF 2014a). However, despite these guidelines and widespread use of routine VL testing in wealthy countries, the cost and complexity of developing a nationally scaled VL monitoring program has made it difficult to implement such programs in LMICs (Roberts et al. 2016, Lecher et al. 2015).

[1] Defined by the World Health Organization as greater than 1000 copies/mL.

1.1.3 Kenya's National HIV Control Programme

In response to Kenya's HIV epidemic, NASCOP was one of the first programs in sub-Saharan Africa to establish an ambitious national VL monitoring approach in 2012. While the VL monitoring program initially only targeted individuals with suspected virological failure, by 2014 NASCOP guidelines echoed the WHO Consolidated Guidelines in recommending routine VL testing every six months for all patients on ART, with a 12-month period recommended between tests after two consecutive tests with VL <1000 copies/mL (NASCOP 2015). This was a significant undertaking given the estimated population of 1.6 million HIV-positive Kenyans (UNAIDS 2016). The program was designed to partner public health facilities with the national laboratory system, all of which is monitored by the NASCOP health management information system and dashboard. In 2017, NASCOP processed 1.04 million VL tests, with an average of 86,727 VL tests processed on a monthly basis (NASCOP 2018). An average of 67 percent of these tests use plasma or EDTA whole blood, and 33 percent use dry blood spots (DBS) (NASCOP 2018). Eighty-four percent of VL test results in 2017 indicated VL suppression (NASCOP 2018).

Laboratory testing in Kenya is typically performed at the 10 Level 4 national reference labs, which have required an upfront investment in infrastructure and equipment. The number of clinical sites networked to these national labs offering VL testing increased from 722 sites in 218 districts in 2012 to approximately 2,000 sites in more than 300 districts by 2016 (Mwau et al. 2018). Reference labs equipped with large, automated, and high throughput diagnostic instruments can process hundreds of batched samples each day. The central lab approach to VL testing relies on a complex transport network and medical infrastructure to collect samples from hospitals and clinics country-wide, and it requires sophisticated tracking mechanisms to ensure results are returned to patients in a timely manner. Most methods of laboratory testing, including the machines evaluated in this study, require venous blood collection (plasma), cold chain, and management of blood samples by trained personnel (UNITAID 2015).

In Kenya, national laboratories use Abbott Laboratories' RealTime m2000sp/m200rt systems, Roche Diagnostics' COBAS® AmpliPrep/COBAS® TaqMan®, and similar systems. These platforms were developed to handle high testing volumes and are sensitive enough to detect very low limits of VL counts. These platforms are considered state-of-the-art for VL monitoring in both high-and low-resource settings (Figure 1).

Figure 1: Central Laboratory Equipment (The Global Fund 2017)

Abbott Laboratories RealTime m2000sp/m2000rt system

- RT HIV-1 qual & quant tests
- 40 copies/ml limit of detection
- 96-144 tests per 8 hours
- Input sample 1ml plasma or DBS
- Full lab capacity required
- Global Fund agreement pricing low as $10.50 for reagents per test available*
- Regulatory status – CE-mark for VL and EID

Image from Abbott.com

*Volume and term commitments apply for reagent rental program; other costs excluded

Roche Diagnostics COBAS®AmpliPrep/COBAS® TaqMan® system

- CAP/CTM HIV-1 v2.0 qual & quant tests
- 20 copies/ml limit of detection
- 144 tests per 8 hours
- Input sample 1ml plasma
- Full lab capacity required
- Global Access Program pricing low as $9.40 for reagents + consumables per test available*
- Regulatory status – CE-IVD for VL and EID

Image from Roche.com

*Volume and term commitments apply for reagent rental program; other costs excluded

However, numerous challenges with central labs have been documented, including the need for expensive equipment and maintenance contracts, staffing shortages of specialized technicians, laboratory workflow and sample transport delays, and accuracy of results reporting (Fox et al. 2016, Stevens et al. 2014, Lecher et al. 2015, Roberts et al. 2016). While NASCOP lists the national turnaround time (TAT) for 2017 as 14 days, primary data collection done in Kenya as part of a related study by the USAID-funded Health Finance and Governance (HFG) project suggests that processing samples in laboratories can lead to longer delays in receipt of test results for some patients. These delays in test result receipt may also result in delays in switching to a more effective line of treatment when the current regimen is failing. A study using NASCOP data from 2018 found that the longest processing times appeared to be the time between lab receipt of the VL sample and lab processing of the sample, followed by transport time to the lab (Mwau et al. 2018). Based on our primary data, another delay can lie in the actionable receipt of the test results. Laboratory-based VL testing requires two clinical visits, one for a blood draw and one to review results. For clients with frequent clinical visits, results can be reviewed at the following appointment in 1–2 months. However, in differentiated HIV care, clients perceived to be stable may only have appointments every 6–12 months, creating a long delay in test result receipt, unless additional attempts are made to reach clients.

Delayed treatment switch following treatment failure has been associated with lower rates of virological suppression on second line therapy, increased risk of opportunistic infections, and increased odds of mortality in low-resource sub-Saharan African settings (Levison et al. 2011, Petersen et al. 2014, Calmy et al. 2007). Further, patients receiving their VL results are significantly more likely to be retained in care and their knowledge of VL values may help improve adherence to therapy, although randomized evidence is limited (Brown et al. 2016, Bonner et al. 2013). Another study set in South Africa indicated that lengthy (6–12 month) delays between virological failure detection and switching from first to second line ART treatment were due to difficulty receiving test results; however, these delays were not found to be associated with earlier virological failure on second line ART (Narainsamy and Mahomed 2017).

Cost is another consideration for NASCOP. All VL testing and counseling is offered free of charge to clients at public facilities, meaning that NASCOP shoulders a significant financial commitment. In 2014, the UNAIDS Diagnostics Access Initiative announced a Global Access Program partnership with Roche Pharmaceuticals to establish a price ceiling of USD $9.40 per VL test in 83 LMICs, representing a 40 percent cost reduction (UNAIDS 2014b). The following year, the Global Fund to Fight AIDS, Tuberculosis and Malaria announced framework agreements with seven diagnostics manufacturers which aimed to save $30 million over three years and improve the transparency and competitiveness of the market (Global Fund 2015). Point-of-care (POC) VL testing technologies may lower costs further, and unit costs are also expected to decrease as volume increases (UNAIDS 2016).

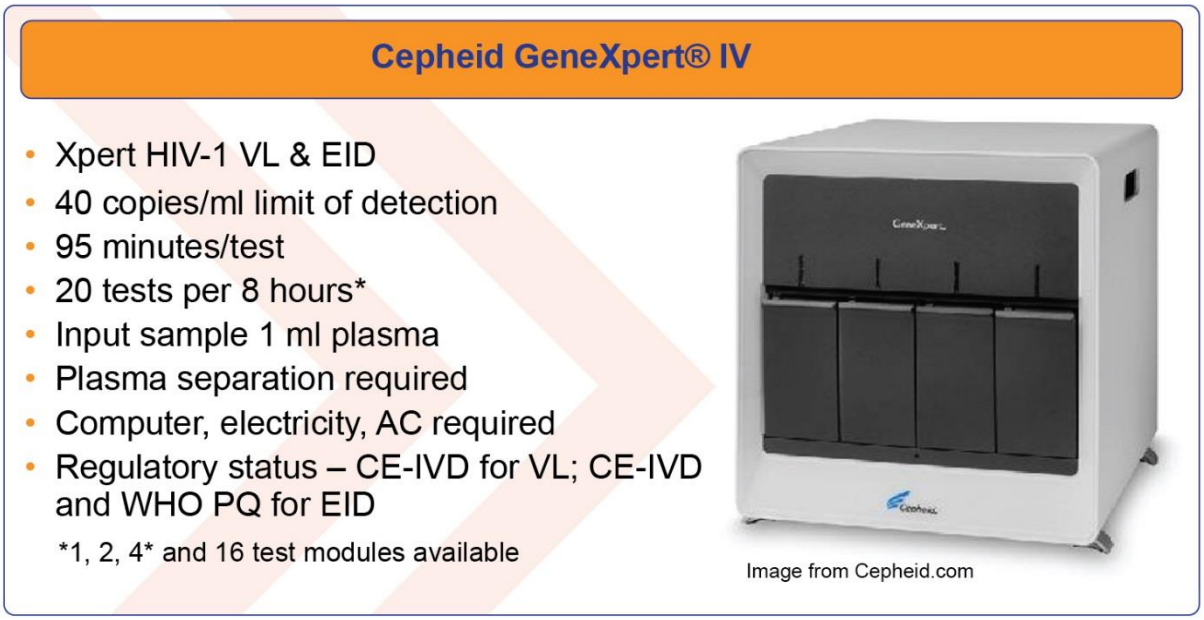
As reported in a related study, costs of high-capacity, lab-based testing platforms from Abbott, Roche, bioMérieux, and Siemens range from $40,000 to over $200,000, while costs per test from the same companies' platforms have ranged between $20 and $90 across various developing country contexts (Cintron et al. 2017, Murtagh 2013). Despite the costs, public sector VL testing programs have recently begun or expanded in several sub-Saharan African countries. Médecins Sans Frontières (MSF) reported VL unit costs from national- and district-level public sector programs in 2013: $43.42 in Kenya, $39.03 in Zimbabwe, $35.38 in Malawi, $34.17 in Lesotho, and $24.90 in Swaziland. Reagents and consumables accounted for an average of 63 percent of costs in these programs (MSF 2013).

1.1.4 Point-of-Care Testing: An alternative to laboratory monitoring

POC VL testing is a novel approach to monitoring HIV treatment and adherence in low-resource settings. POC testing has been championed as a promising alternative to central laboratory VL testing by the WHO, UNITAID (Market catalysts; Geneva), the Bill and Melinda Gates Foundation, the Clinton Foundation, The United States President's Emergency Plan For Aids Relief (PEPFAR), and the African Society of Laboratory Medicine (WHO 2013, Stevens et al. 2014). POC CD4 testing compared to central lab CD4 testing has shown promising impact and cost-effectiveness outcomes in low-resource settings (Hyle et al. 2014, Dorward et al. 2018). While there is not a universally accepted definition of POC testing, typically POC devices process tests individually (as opposed to in batches) and are intended for lower patient testing loads (4–20 samples per day) (UNITAID 2015). POC tests are practical in low-resource settings because they can function without constant electricity or running water and they do not require specialized technicians. No transport and minimal infrastructure and maintenance is needed, and test processing can take from 60 to 120 minutes (UNITAID 2015). In 2017, the WHO approved the first quantitative POC HIV VL assay for use in low-resource settings (WHO 2017). POC VL testing is promising because it has the potential to improve patient-centered care and improve the timeliness of client receipt of test results.

POC equipment such as the Cepheid GeneXpert® IV may support improved, decentralized VL testing that circumvents some of challenges associated with laboratory testing. Several similar POC VL assays have also been validated in decentralized clinics in southern Africa, including the Alere q NAT (Alere, Waltham, MA, US), SAMBA I/II semi-Q (Diagnostics for the Real World Ltd., Cambridge, UK), and Liat HIV Quant (Roche Diagnostics, Basel, Switzerland) (Dorward et al. 2018). However, the widespread use of the GeneXpert platform for tuberculosis diagnostics may favor its selection over its competitors. Key information on select POC equipment used in this model are included in Figure 2.

Figure 2: Point-of-Care Equipment (The Global Fund 2017)



Alere™ q

- Alere q HIV-1/2 Detect assay
- 1,000 copies/ml limit of detection
- 50 minutes/test
- 8 tests per 8 hours
- Input sample 25 ul whole blood
- Fully contained reagent cartridge
- Limited battery-powered operability
- Regulatory status – WHO PQ & CE-IVD for EID; VL in progress

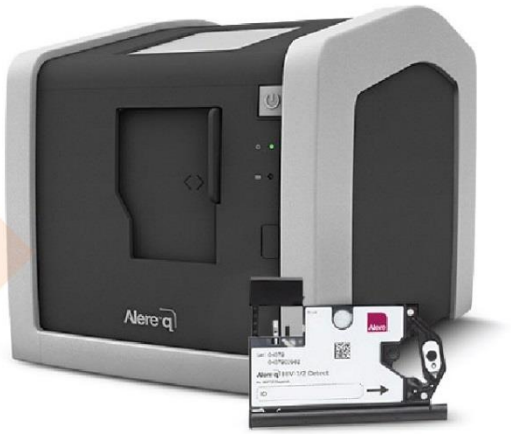


Image from Alere.com

1.2 Objective

The HFG project was tasked with assessing the cost-utility of POC VL monitoring of HIV-positive Kenyans on ART compared to the current laboratory VL monitoring approach. To our knowledge, there have not been any studies comparing the cost-effectiveness of POC VL testing to laboratory testing for routine monitoring of HIV treatment at scale from the health system perspective in Kenya, or in any low-resource setting. There is significant political momentum globally to transition to POC VL testing given the logistical challenges presented by laboratory monitoring. However, the costs and impact of this transition are unclear because there is limited empirical data on implementation. Therefore, this question can only be answered at this point using a hypothetical modeling approach.

Using costing data and empirical data collected for a related study as well as data from the literature, this study developed a Markov model to conduct a cost-effectiveness analysis of central vs. POC approaches to VL monitoring at scale from the health system perspective in Kenya. This report provides the Kenyan Ministry of Health and the global community with recommendations for best practices on scaling up VL testing in low-resource settings.

2. METHODOLOGY

2.1 Model Overview

This study compares the costs and outcomes of the provision of VL monitoring at scale to Kenyans on ART through a laboratory (status quo) or POC approach. This analysis takes the perspective of the Kenyan Ministry of Health when calculating costs. A Markov model was developed to reflect the movement of a hypothetical cohort of one million ART patients through possible health states. The hypothetical cohort consists of Kenyans of average age, sex, and co-morbidities participating in VL monitoring. All members of the cohort start on first line ART (FLART) and then progress through different states. The time horizon for this analysis is 10 years with a lifetime analytic time horizon for disability-adjusted life years (DALYs). The 10-year time horizon was selected over a lifetime time horizon because it is more applicable to national budget cycles and program funding. Six-month cycles were used to mirror the WHO-recommended VL testing intervals.

In this study, POC technology is represented by the Alere™ q and Cepheid GeneXpert® IV testing platforms. Representative tests for central lab include Abbott Laboratories RealTime m2000sp/m200rt system and the Roche Diagnostics COBAS® AmpliPrep/COBAS® TaqMan® system.

2.1.1 Model Assumptions

Several assumptions are built into this model, which we describe here.

Cohort demographics and treatment

Cohort demographics are assumed to mirror adult age, gender, and co-morbidity distributions found in the population currently being served by NASCOP. In a recent study of FLART patients in Kenya, treatment failure was not significantly associated with age, gender, or time on ART, though the sample may have been too small for detection (Brooks et al. 2016). While individual demographics can significantly impact health outcomes, at the scale of this intervention, demographics were assumed to average out and did not affect participant flow through the model.

While all cohort participants begin on FLART in the model, they are not all initiating FLART. They are initiating VL monitoring of their FLART. The model does not therefore differentiate between those who have been on treatment for years and those who are initiating treatment.

No third line ART

Third line drugs typically cost between six and 14 times more than first and second line therapies (Carter 2017). Affording third line drug regimens is a challenge for treatment programs in low- and middle-income settings, where budgets are limited and need is much greater for first and second line treatments. As one example, Brazil currently spends 40 percent of its ART budget on the 5 percent of patients who are in need of third line treatment (Carter 2017). While the NASCOP has developed a 2016 Toolkit for Third Line Antiretroviral Therapy for Service Providers in Kenya, the program acknowledges that patients failing second line ART have limited remaining options (NASCOP 2016). For these reasons, a third line treatment was not included in the model.

Comparable sensitivity and specificity

Key differences between central lab and POC VL testing were initially hypothesized to include the sensitivity and specificity of the test results. However, at the 1000 copies/mL threshold recommended by the WHO to determine whether a patient was failing treatment, sensitivity and specificities were quite high for both tests and significant differences between them have not been established (Sollis et al. 2014, Garrett et al. 2016, Jani et al. 2016). Both approaches have near-perfect specificities, making them comparable in this regard. The Xpert HIV-1 VL (Cepheid, Sunnyvale, US) detects virological failure (>1000 copies per mL) with 94 percent sensitivity and 99 percent specificity (Dorward et al. 2018). The COBAS® AmpliPrep/COBAS® TaqMan® HIV-1 Test has a specificity of 100 percent and an analytical sensitivity of ≥95 percent (Roche 2018). Both approaches have similar capacities to detect high and near-zero VL levels.

Plasma vs. DBS testing

This study only considered POC tests that used plasma or whole blood, rather than DBS testing. DBS is a promising next generation of POC testing but it currently has sub-optimal sensitivities and specificities compared to plasma testing (Pannus et al. 2016).

WHO and NASCOP guidelines

WHO and NASCOP guidelines and definitions were followed for this model. Viral suppression is defined as <1000 copies/mL. Virological failure is defined as >1000copies/mL. Enhanced adherence counseling (EAC) is provided for those receiving test results indicating >1000 copies/mL. ART switching criteria follows WHO guidelines.

Loss to follow-up

Once a patient enters the lost to follow-up (LTFU) state, the patient does not rejoin the system. In reality, many of these patients may rejoin care in the future but follow-up data on treatment pathways and health outcomes is limited for these patients.

Delays

Test result delays assumed to be clinically relevant in this model are delays that are six months or longer. While a larger number of patients may experience a shorter delay (<6 months) in receiving their test results, longer delays are more clinically relevant for patients failing ART, secondary transmission, and LTFU. Further, six-month delays conform to the structure and cycle-length used for the model. Even if a patient receives his/her test result at a following appointment six months later and the result indicated that the patient was failing treatment (VL>1000 copies/mL), the patient would still need to receive EAC to rule out the possibility that failure was due to behavior, rather than biological failure for that line of treatment before moving to the next treatment line. This means that they would spend another six months on the same line of treatment.

Mortality

Mortality accounted for in this model is all-cause mortality, with the assumption that deaths not related to the VL monitoring approach will not be differentially affected by VL monitoring approach, and therefore not impact the relative cost-utility of the results (Estill et al. 2013, Salazar et al. 2014).

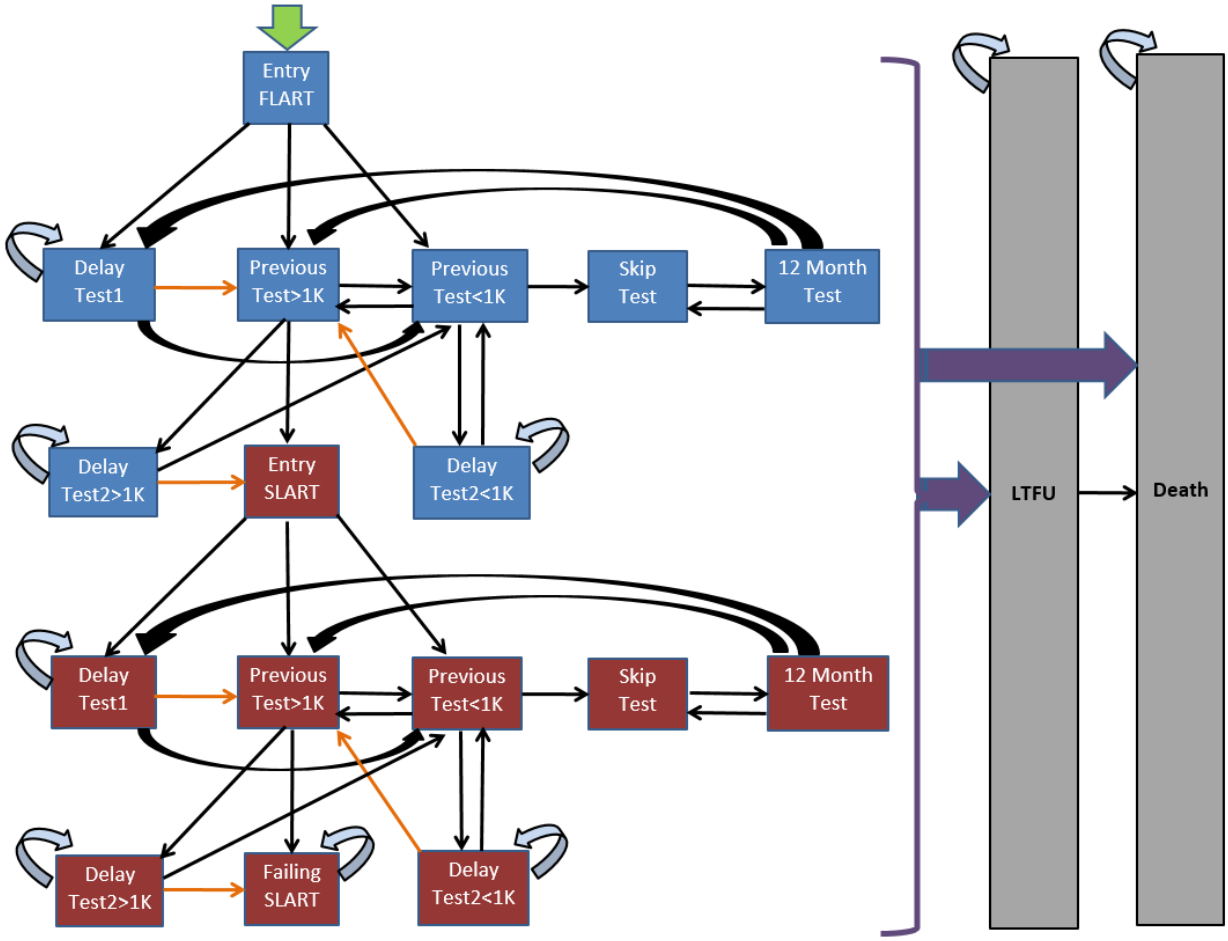
Uncontrolled viral load

Some individuals who experienced a delay in receiving their test results would be failing treatment (VL>1000 copies/mL). Because these individuals would not receive EAC or move to a second line of treatment, they would have uncontrolled VL for six months and be more likely to transmit HIV to a partner.

2.1.2 Model Narrative

The Markov model in Figure 3 depicts the possible states and flow patterns through which the cohort moves over the 10-year time horizon.

Figure 3: Markov Model Diagram



FLART – First Line Antiretroviral Therapy
SLART – Second Line Antiretroviral Therapy
LTFU – Lost to follow-up
>1K – Previous VL test result was higher than 1000 copies/mL
<1K – Previous VL test result was lower than 1000 copies/mL
Orange arrows indicate individuals who may have had uncontrolled VL during delay

All individuals enter the model at the *Entry FLART* state (Figure 3). States in blue indicate individuals who are on FLART. States in red indicate individuals who are on second line ART (SLART). In each state, in addition to the transitions described below, there is a risk of LTFU and death. At the end of the first six-month cycle in the *Entry FLART* state, all individuals in *Entry FLART* who are not LTFU or deceased take a VL test.

- If the individual received the test result and the result indicated more than 1000 copies of the virus, s/he moves to the FLART *Previous Test >1K* state where s/he receives EAC.
- If the individual received the test result and the result indicated less than 1000 copies of the virus, s/he moves to the FLART *Previous Test<1K* state.
- If s/he does not receive the test results, s/he moves to the blue (FLART) *Delay Test1* state and does not receive EAC.

After six months (1 cycle) in the FLART *Previous Test>1K* state, individuals who have not been LTFU or deceased take a second VL test.

- If the test indicates that they are failing treatment (>1000 copies/mL), then the individual moves to the *Entry SLART* state where they initiate SLART. This is timely and appropriate treatment.
- If the test indicates that they are not failing treatment (<1000 copies/mL), they move to the *Previous Test<1K* state.
- If they do not receive their tests, they move to the blue *Delay Test2>1K* state.

After six months (1 cycle) in the *FLART Previous Test<1K* state, individuals who have not been LTFU or deceased take a second VL test.

- If the test indicates that they are failing treatment (>1000 copies/mL), then the individual moves to the *FLART Previous Test>1K* state.
- If the test indicates that they are not failing treatment (<1000 copies/mL), they move to the *Skip Test Holding* state, because they have received two consecutive tests indicating that they are not failing treatment.
- If they do not receive their test, they move to the *FLART Delay Test2<1K* state.

After six months (1 cycle) in the *FLART Delay Test1* state, individuals who have not been LTFU or deceased may or may not receive their test result. Regardless, the test results are old and so everyone takes a new VL test.

- If the test indicates that they are failing treatment (>1000 copies/mL), then the individual moves to the *FLART Previous Test>1K* state.
- If the test indicates that they are not failing treatment (<1000 copies/mL), they move to the FLART *Previous Test<1K* state.
- If they do not receive their test again, they remain in the *FLART Delay Test1* state.

Everyone in the *Delay Test 2>1K* state has previously had a test indicating that they are failing treatment. After six months (1 cycle) in the *Delay Test2>1K* state, individuals who have not been LTFU or deceased take another VL test.

- If the test again indicates that they are failing treatment (>1000 copies/mL), they move to the *Entry SLART* state.
- If the test indicates that they are not failing treatment (<1000 copies/mL), they move to the *FLART Previous Test<1K* state.
- If they do not receive their test again, they remain in the *Delay Test 2>1K* state.

Everyone in the *Delay Test 2<1K* state has previously had a test indicating that they are not failing treatment. After six months (1 cycle) in the *Delay Test2<1K* state, individuals who have not been LTFU or deceased take another VL test.

- If the test indicates that they are failing treatment (>1000 copies/mL), they move to the *Previous Test>1K* state.
- If the test indicates that they are not failing treatment (<1000 copies/mL), they move to the *Previous Test<1K* state. Even though they have had two tests indicating that they are not failing treatment, the tests did not occur within the past 12 months and so the individual cannot move to the *Skip Test* state.
- If they do not receive their test again, they remain in the *Delay Test 2<1K* state.

The *Skip Test* state was included in the model to acknowledge the WHO guidelines, which recommend that, after two consecutive VL tests within 12 months which indicate that an individual is not failing treatment, a 12-month period of no VL testing is warranted. After six months (1 cycle) in the *Skip Test* state, all individuals in that state who are not LTFU or deceased move to the *12 Month Test* state without taking a VL test.

After six months (1 cycle) in the *12 Month Test* state, all individuals who are not LTFU or deceased take a VL test.

- If the test indicates that they are failing treatment (>1000 copies/mL), they move to the *Previous Test>1K* state.
- If the test indicates that they are not failing treatment (<1000 copies/mL), they move to the *Skip Test Holding* state.
- If they do not receive their test, they move to the *Delay Test1* state.

In the *LTFU* state, individuals can either remain LTFU or move to the *Death* state but they cannot return to any other states in the model. The *Death* state is an absorbing state.

The SLART states follow the same pattern as the FLART states. Because there is typically not a third line ART available in sub-Saharan Africa, if an individual is failing SLART (indicated by two consecutive positive VL tests), then they will move to the *Failing SLART* state and from there will move to either *LTFU* or *Death* states.

2.2 Transition Parameters

Transition parameters shape how the cohort flows through different states in the model over time. This model includes a number of transition parameters that are presented in Table 1, and will be discussed in the following sections.

Table 1: Transition Parameters

Monitoring Model	Treatment Line	Probabilities of Occurence per 6-month Cycle	Value Estimate	Min	Max	Source
POC	FL/SL	Delay in test receipt	0	0	0	Assumption
Laboratory	FL/SL	6+ month delay in test receipt	0.022	0.011	0.081	NASCOP 2018; HFG primary data collection
Test agnostic	FL/SL	VL test result >1000 copies/mL	0.165	0.133	0.186	NASCOP 2018; Mwau et al. 2018
Test agnostic	FL/SL	Resuppression after VL test result of >1000 copies/mL with EAC	0.600			Billioux et al. 2015
Laboratory	FL/SL	Resuppression after VL test result of >1000 copies/mL without EAC	0.246			Billioux et al. 2015; Chung et al. 2011
Test agnostic	FL/SL	General LTFU	0.021	0.01	0.05	Ayah 2018, Arnesen et al. 2017, Mberi et al. 2015
Laboratory	FL/SL	LTFU during delay	0.050			Arnesen et al. 2017, Mberi et al. 2015
Test agnostic	FL/SL	Death (all cause)	0.019	0.009	0.027	Biset Ayalew 2017, Rubaihayo et al. 2015,
Test agnostic	SL	LTFU while failing SLART	0.100			Assumption
Test agnostic	SL	Death while failing SLART	0.048			Pujades-Rodriguez 2010
Test agnostic	SL	Death while LTFU	0.033			Average of general death and failing ART death
Test Agnostic	FL/SL	Probability of HIV transmission with uncontrolled VL	0.010	0.005	0.015	Mujugira et al. 2016

FL: First line
SL: Second line

2.2.1 NASCOP Delay in Test Receipt

The primary hypothesized difference in POC and laboratory testing other than costs is the delay in receipt of test results. When considering the clinical impact of delay in test receipt, it was important to consider both the probability of delay, and the length of that delay. There are 10 laboratories serving 2,198 facilities that received a total of 1,093,527 samples for VL testing in 2017. During this time, the probability of a laboratory rejecting a sample after receipt was 0.69 percent, with a range of rejection probabilities from 0 percent to 1.4 percent depending on the laboratory (NASCOP 2018). After preliminary testing, 1.47 percent of test results were inconclusive and required a redraw of blood from the patient. The probability of required redraw ranged from 0.4 percent of received samples to 6.8 percent of received samples, depending on the facility. This indicates that there was an average probability of delay of 0.022 across laboratories, a figure that does not include samples that were lost or damaged in transit to the laboratory and results that did not reach their recipients after a conclusive test result. This is captured as the point estimate for the delay transition parameter. The length of these delays is not known but since patients would be required to coordinate timing and transport to revisit the facilities to provide a new blood sample, and then return for results, it is probable that these delays could last a number of months. The lower bound for the delay parameter was estimated to be a 50 percent decrease in probability from the average delay (0.011). The upper bound (0.081) was derived from primary data collection on TAT and will be discussed in the next section.

NASCOP reports on TATs as the number of days between the date the sample is collected from the patient and the date that the test result is received by the health facility, rather that the date the test is received by the patient. In 2017, the average TAT for VL samples that were able to be processed was 14 days, with a range from seven days to 25 days depending on the laboratory (NASCOP 2018). However, NASCOP does not publish data on every TAT and therefore there is limited information on the probability of a long (5+ month) TAT which is an important component of this model. Delays of six months or longer not only fit better with the WHO-driven cycles in this model, but also are more clinically relevant for patients failing ART, secondary transmission, and LTFU. Primary data on TAT was collected to better understand the probabilities of these longer delays and complement the NASCOP data by recording the point at which patients receive test results (see next section).

2.2.2 Turnaround Time Primary Data Collection

To capture additional data on VL test TATs, clinical and laboratory records data were collected from level 2 and 3 health facilities in Siaya County, Kenya, and the Kenya Medical Research Institute (KEMRI) lab in neighboring Kisumu County. Out of 20 facilities selected using probability proportional to size sampling during initial data collection on VL testing costs in a parallel costing study, 15 facilities were purposively selected on the basis of patient volume for additional sampling of TAT data. Two pilot facilities were also selected and visited first.

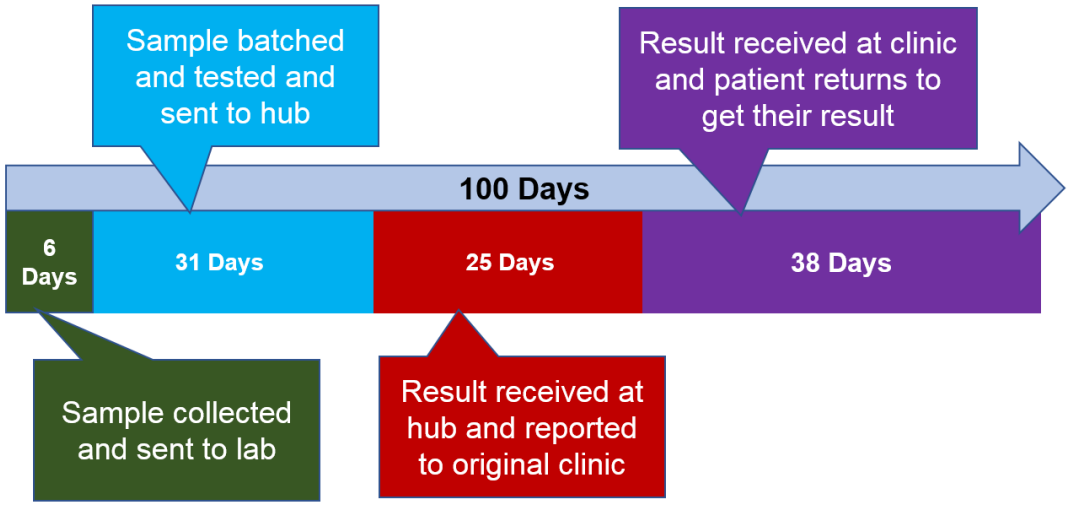
The research protocol used in the original costing study received ethical review approval through the Abt Associates Institutional Review Board, the KEMRI Ethical Review Committee and Scientific Steering Committee, and the Centers for Disease Control and Prevention (CDC) Institutional Review Board. The KEMRI Ethical Review Committee and Scientific Steering Committee also approved the collection of TAT data by KEMRI lab staff. All review of sample receipt logs and result records was conducted by health facility staff, and no identifiable data were collected.

Data were extracted at facilities from sample receipt logs and then confirmed by facility lab staff, by clinical staff, and against KEMRI records when possible. VL sample TATs were collected for approximately 20 adult (age>17) patients on ART and receiving VL monitoring at each facility (n=295). To capture the distribution of TATs throughout 2016, the first patient record selected was a patient who had a blood sample drawn in January, the second was a patient with a sample drawn in February,

and so on until all months were covered and the process started over. If no patients had samples collected in a given month, the month was skipped.

Collected data points included: age of patient, patient residence region, date of ART initiation, number of patient visits in the last 12 months, date VL sample was taken, date sample arrived at hub hospital, date sample arrived at KEMRI lab, date VL result arrived at hub hospital, date result arrived at origin facility, date patient received result, VL test result, reason for testing, and a delay narrative, if available. Staff at all facilities were also informally interviewed regarding delays in the testing cycle. Of the 295 patient records sampled, 16.6 percent had missing values for the total time (in days) between the date the sample was taken and the date the patient received the VL test result. Of the remaining 246 valid results, 8.13 percent of patients (n=20) did not receive their test results for at least six months and an additional 12 patients did not receive their test results for at least five months (cum. 13.01%). On average, patients received their results 94 days after submitting a sample. Results arrived at the clinics an average of 58 days after the sample was taken, indicating that one of the largest delay components was between clinic receipt and patient notification. However, there was not a significant difference in notification times between suppressed and unsuppressed results ($p > 0.2$). While this was a small sample in one region, the substantial differences in TAT (including sub-phases) from national data warrant further investigation of TAT in this region and reconsideration of how TAT is defined and reported by NASCOP.

Figure 4. Average Turnaround Timeline from Primary Data Collection



2.2.3 Adherence, Treatment Failure, and Resuppression

In 2017, NASCOP successfully performed 1,040,726 VL tests, of which 83.55 percent demonstrated VL suppression (NASCOP 2018). This means that 16.45 percent of test results indicated VL higher than 1000 copies/mL. NASCOP data does not differentiate test results by first or second treatment lines or by time on ART but both first and second lines of ART have demonstrated high success rates for viral suppression and therefore no differentiation was made in model parameters for first and second line ART probabilities of suppression (Kityo et al. 2014, Boender et al. 2016). Further supporting the decision not to differentiate between probabilities of failure between first and second line treatments is the fact that virological failure rates can vary significantly, even within treatment lines. A systematic review and meta-analysis of 19 studies reporting outcomes of second line ART in resource-limited settings indicated that the proportion of adult patients experiencing virological failure varied widely, from 8.3 percent to 41.2 percent, at 24 months (Ajose et al. 2012).

Non-adherence to a treatment regimen is a well-established reason for high VL among patients on ART (MSF 2016). After receiving a VL test result of >1000 copies/mL, patients are offered EAC per NASCOP guidelines (NASCOP 2016). The goal of EAC is to review and address psychological, emotional, and socio-economic factors that may contribute to a patient's poor adherence in a nonjudgmental way. Three sessions are recommended, but not always achieved. Evidence linking EAC to adherence, suppression, resuppression, and retention in care is limited, but EAC is still recommended in most national and international HIV treatment guidelines (Jobanputra et al. 2015; Billioux et al. 2015, Chung et al. 2011, Langebeek et al. 2014). One randomized control trial in Kenya found that patients receiving adherence counseling were 29 percent less likely to experience poor adherence compared to those who received no counseling, and patients receiving intensive early adherence counseling were 59 percent less likely to experience viral failure (Chung et al. 2011). Resuppression is the achievement of a VL test result of <1000 copies/mL after receiving a previous test result of >1000 copies/mL. The probability of resuppression after adherence counseling, confirmed by a second VL test, is variable, and can range widely depending on the setting (MSF 2016). In this model, resuppression after receiving a positive test is parameterized at a probability of 0.60 whereas resuppression after delayed receipt of a positive test is 59 percent lower at 0.246 (Billioux et al. 2015, Chung et al. 2011).

2.2.4 Loss to follow-up, secondary transmission, and death

Rates of LTFU among patients on ART in sub-Saharan Africa range widely depending on how LTFU is defined and a number of demographic and other contributing factors (Arnesen et al. 2017, Fox et al. 2016, Reidy et al. 2014, Mberi et al. 2015, Ayah 2018). Some studies suggest that rates of LTFU are often overestimated due to patient transfer to other facilities and inadequate medical records (Fox et al. 2016, Yehia et al. 2015, Geng et al. 2011). A recently published study based in Kenya found that over a 12-month study period 4.2 percent of patients were LTFU, while two recent studies in South Africa noted higher rates of LTFU that convert into a six-month transition rate of .051 (Ayah 2018, Arnesen et al. 2017, Mberi et al. 2015). There is limited evidence on the impact of delayed test results on LTFU; however, it is assumed to be higher than for those who are more informed and engaged in care. Therefore, the high value (0.051) in the standard LTFU range is used for LTFU from delay in this model.

One of the key goals of the 90-90-90 approach championed by UNAIDS is to reduce the risk of secondary transmission by ensuring VL suppression. The landmark randomized controlled trial in Uganda that solidified ART as a preventative measure against HIV transmission was HPTN 052, which demonstrated a 93 percent reduction of HIV transmission within sero-discordant couples when comparing the group in which the HIV-infected partner was assigned to early ART with the group in which the HIV-infected partners was assigned to the delayed ART group (Safren et al. 2015). Incidence of HIV transmission to the uninfected partners was 2.08 infections per 100 person years in the period before the HIV-positive partner started ART and incidence fell to zero after six months of ART (Mujugira et al. 2016). Delaying treatment switch for patients with unsuppressed VL therefore increases the risk of sexual transmission of HIV. This model tracks uncontrolled VL among individuals who are unsuppressed (VL>1000 copies/mL) due to delay and LTFU. While the total number of secondary transmissions is included as a secondary outcome of the model, the costs and outcomes of those infected were not incorporated into the model as they were not part of the original cohort. This modeling could be done as part of larger population-based model.

The introduction of ART has significantly reduced mortality and increased life expectancy globally for people living with HIV, including in low-resource settings (Rubaihayo et al. 2015). A *Morbidity and Mortality Weekly Report* from 2013 presented ART outcomes for six African countries between 2004 and 2012; the most successful program was in Uganda with 91 percent of men and 94 percent of women alive and still enrolled on therapy after six months (Ettiègne-Traoré et al. 2013). A recent systematic

review of mortality in ART patients in Ethiopia found a range of mortality incidence densities from 0.2 to 10.74 per 100 person-years, with a majority of the included studies reporting an incidence density of 1.89–5.3 per 100 person-years (Biset Ayalew 2017). This was similar to an earlier Rwandan study that reported a mortality incidence of 3.7 per 100 person-years and a Ugandan study that reported mean annual mortality percentages of 1–4 (Geng et al. 2011, Rubaihayo et al. 2015). Studies on SLART mortality outcomes in low-resource settings are limited and it is difficult to compare SLART outcomes to FLART outcomes due to survivor bias and other confounding factors. However, SLART is generally considered as effective as FLART, with overall low rates of attrition and death and moderately high rates of virological suppression (Shearer et al. 2016). Further, studies have shown that more deaths happen earlier in treatment, again potentially due to survivor bias or experience than treatment regimen (Gunda et al. 2017). Therefore, the lower end of the range of death probabilities was used for SLART patients.

Probability of death while LTFU and while failing SLART was derived from older studies estimating mortality in untreated populations. One study estimated that the five-year survival probability for patients without highly active ART with CD4 counts of 200–3500 cells per mm3 was 52.7 percent (CI: 15.7%–80.2%) (Zwahlen and Egger 2006).

2.3 Costs

2.3.1 Collecting Costing Data

Costs for laboratory VL testing are sourced from a prior HFG costing study of a large testing network that includes the KEMRI/CDC lab in Kisumu City, Kenya, and a probability sample of 21 health facilities in neighboring Siaya County (Cintron et al. 2017). The facilities send whole blood or DBS samples to hub hospitals for plasma separation, which then sends plasma samples to KEMRI for testing. Test results are sent electronically from KEMRI back to the hub hospitals, from which paper copies of the results are sent back to the health facilities. The unit cost of a VL test includes all supplies, equipment, human resources, training, overhead, and reagents required to test a sample, throughout the processes of sample collection, centrifuging, and testing, along with costs for quality assurance and sample transportation, which is conducted by private couriers who charge a fee per batch delivered. Kenya uses two machines, the Abbott Laboratories RealTime m2000sp/m200rt system and the Roche Diagnostics COBAS® AmpliPrep/COBAS® TaqMan® system, procured through a reagent rental program where equipment, maintenance, and machine-specific training costs are all included in the negotiated cost of reagents. The reported laboratory VL unit cost is the average of the Abbott and Roche-specific unit costs, as testing is split between the two platforms.

POC VL testing costs include the same components as central lab costs, excluding transportation, but are based on secondary data and estimation as well as the aforementioned costing study. Two platforms are being considered for use in Kenya, the Alere™ q and the Cepheid GeneXpert® IV system, and the reported unit cost is the average of costs estimated for each platform. Unlike the Abbott and Roche machines used at the central lab, reagent rental agreements from Alere and Cepheid are not currently available and their equipment must be purchased. Reagent and equipment costs for each POC platform are dependent on procurement agreements between countries or regions and manufacturers; therefore, the costs used here are approximations based on figures reported by the Global Fund (2017) and MSF (2014). The cost of POC platforms per test is calculated as the sum of the machine price and maintenance agreements for an assumed five-year lifespan divided by the maximum number of tests possible over the same time period, assuming eight-hour working days and 250 working days per year. Other supply, equipment, and overhead costs are based on the items and spaces used for sample collection at health facilities, obtained through the costing study. Human resource costs for sample collection are based on the costing activity, and costs of POC testing are based on the assumption that

sample collection and testing will be performed by the same staff, as intended by the platforms ease of use. Training costs are also based on the costing activity, but account for an additional training on the new POC platforms, which manufacturers indicate should take less than a day (UNITAID 2015). Quality assurance costs are calculated as 5 percent of the total unit cost (pre-quality assurance), as that is what we found them to be for central lab testing and protocols for POC quality assurance are not yet defined.

Costs of FLART and SLART per person per six months (Table 2) are based on the WHO Global Price Reporting Mechanism and Kenya's 2016 ART guidelines. FLART consists of annual course of Tenofovir + Lamivudine + Efavirenz [TDF+3TC+EFV], while SLART consists of Zidovudine + Lamivudine [ZDV+3TC] and Atazanavir + Ritonavir [ATV+RTV]. Use of generic versions of each drug are assumed. Non-ART costs of FLART and SLART are not included, as we assume they would not differ by VL testing approach.

Table 2: Component Costs for Laboratory and POC VL Testing

Cost Item	Value	Min	Max	Source
FLART per person per 6 months	$42.52			WHO 2016
SLART per person per 6 months	$125.96			WHO 2016
POC per test	**$29.74**	**$26.16**	**$33.33**	Cintron et al. 2017
POC machine	$2.08			
POC supplies/reagents per test	$23.68			
POC labor for per test	$1.38			
POC overhead per test	$2.60			
Central lab per test	**$24.63**	**$23.92**	**$26.05**	Cintron et al. 2017
Central lab machine	$0.00			
Central lab supplies/reagents per test	$19.00			
Central lab overhead	$3.74			
Central lab labor	$1.34			
Central lab transport	$0.55			
End of Life care	$53.94			Hamers et al. 2012

Unit costs of POC and lab-based VL testing are $29.74 and $24.63, respectively. To the best of our knowledge, there are currently no published program-based costs of POC VL testing in sub-Saharan Africa, but the average unit cost of $29.74 is close to costs estimated by the Global Fund (2017) and manufacturers. The unit cost of lab-based testing falls within the range of previously reported VL testing costs in sub-Saharan Africa, including costs from Kenya (MSF 2013, 2014b, 2014c).

2.4 Disability Weights

The DALY is a universal metric that allows economists and health researchers to compare different populations and health conditions over time. DALYs are calculated by summing the years of life lost and years lived with disability. Disability is calculated through the application of disability weights, quantifications of the severity of the disease sequela, to the amount of time spent in a particular health state. Disability weights are measured on a scale of 0 to 1, where 1 is death (least desirable) and 0 is perfect health (most desirable) (WHO 2010). DALYs capture both the impact of an intervention on an individual's length of life as well as the impact on their health-related quality of life.

The Global Burden of Disease (GBD) study by the Institute for Health Metrics and Evaluation (IHME) is one of the leading sources for disability weights. Table 3 shows the disability weights associated with different HIV states (GBD 2013, GBD 2016).

Table 3: Disability Weights for HIV Health States (GBD 2016)

Code	Sequela	Health State Name	Description	Disability Weight (95% CI)
a	Early HIV without anemia	Generic uncomplicated disease: anxiety about diagnosis.	Has a disease diagnosis that causes some worry but minimal interference with daily activities.	0.012 (0.006–0.023)
b	HIV/AIDS: receiving ART	HIV/AIDS: receiving ART	Generic uncomplicated disease. The person takes daily medication that sometimes causes diarrhea.	0.078 (0.052–0.111)
c	Symptomatic HIV without anemia	HIV cases, symptomatic, pre-AIDS	Has weight loss, fatigue, and frequent infections.	0.274 (0.184–0.377)
d	AIDS without ART without anemia	AIDS cases, not receiving ART	Has severe weight loss, weakness, fatigue, cough and fever, and frequent infections, skin rashes, and diarrhea.	0.582 (0.406–0.743)
e	Death	Death	Patient has died.	1.000

Table 4 maps the above disability weights to the health states in the model. As with the transition parameters, no distinction was made between uncomplicated FLART and SLART states for reasons described previously.

This model used a lifetime analytical timeframe for deaths, meaning that when an individual died during the 10-year timeframe, the calculation would consider years of life lost beyond the 10-year horizon. Lifetime DALYs incurred by death were calculated by multiplying the number of new deaths by an average life expectancy across adults in Kenya. Overall life expectancy in Kenya is 62 years and this was used instead of a specific HIV-positive population life expectancy, because a Uganda study demonstrated that life expectancy for HIV-positive adults starting ART is comparable to life expectancy for all Ugandan adults (Mills 2011, World Bank 2016). Since the cohort ranges in age from 18-62, a life expectancy multiplier of 22 years was used as the average for individuals in cohort. A correction for time spent in the model was included by subtracting the number of years spent in the model from the lifetime DALYs.

Table 4: Disability Weights for Model Health States

State(s)	GBD Disability Weight Code	Description	Disability Weight
FLART	b	All participants in FLART states receive the same weight	0.078
SLART	b	All participants in uncomplicated SLART states receive the same weight	0.078
Failing SLART	d	ART is ineffective and VL exceeds 1000 copies/mL	0.582
Delay	a+b	Patient experiences extra anxiety due to long delay in test receipt	0.090
LTFU	c	Patients have discontinued care and may be experiencing symptoms	0.274
Death	e	Patient has died	1.0

2.5 Discounting

The process of discounting was developed to capture the concept that costs and benefits incurred today are more highly valued than costs and benefits occurring in the future (NICE 2017). Discounting costs reflects preferences for costs to be experienced in the future rather than the present. Discounting health benefits reflects preferences for benefits to be experienced sooner rather than later, though this application is subject to some debate. The WHO recommends a discount rate of 3 percent to be applied annually to both costs and benefits, and that practice has been applied in this model, adjusted by the six-month cycle length (WHO 2010). Alternative discount rates of 0 and 11.5 percent were modeled in sensitivity analysis, discussed in the next section.

2.6 Sensitivity Analysis

Sensitivity analysis allows researchers to test the robustness of the model to parameter inputs and assumptions by varying one parameter at a time and documenting outcomes of interest. Four key parameters in this analysis are 1) the cost differential between monitoring approaches; 2) the probability of six-month (or longer) delay for the laboratory model; 3) the increased probability of LTFU during delay; and 4) the probability of transmission for uncontrolled VL. For each of these parameters, we identified minimum and maximum probable values from the literature or increased and decreased the probability by 50 percent. Because we are more interested in the cost differential between the approaches rather than their absolute costs, we took the difference been the highest cost of POC ($33.33) and the lowest cost of laboratory ($23.92) and included it as Value A and took the absolute difference between the lowest cost of POC and the highest cost of laboratory and included it as Value B (Table 5). For the probability of longer delay in test receipt, the Value A input was 0.011, a 50 percent reduction from the 0.022 probability of six-month delay from NASCOP data, and Value B was 0.081, which came from the primary data collection on TAT. For Probability of LTFU during delay, Value A was a 50 percent reduction from the 0.05 probability used in the base case and Value B was 0.075, a 50 percent increase from the base case. Finally, because discounting outcomes is considered controversial, we also conducted sensitivity analyses where discounting of outcomes (only) was eliminated and where the model applied the Central Bank of Kenya 2016 discount rate of 11.5 percent to outcomes (CIA World Factbook 2016).

Table 5: Sensitivity Analysis Parameter Values and Distributions

Parameter	Value A	Value B	Distribution*
Extreme cost differentials between monitoring approaches ($POC-$Lab)	$9.41	$0.11	Truncated normal
Probability of 6-month delay in test receipt for laboratory	0.011	0.081	Uniform
Probability of LTFU during delay	0.025	0.075	Normal
Discount rate for model outcomes	0.115	0.000	N/A
Probability of transmission due to uncontrolled VL**	0.005	0.015	Normal

**Only used for probabilistic uncertainty analysis.*
***Only used for transmission outcome.*

Probabilistic uncertainty analysis allows researchers to account for substantial uncertainty in a model by assessing the impact of changes in multiple parameters simultaneously, based on their distributions. Using Oracle Crystal Ball, transition probabilities were sampled from normal and uniform distributions and costs of POC and central laboratory tests were sampled from truncated normal distributions (Briggs et al. 2006). Parameters for sampling distributions were derived from point estimates for each variable. For costs, base case values were set as the mean and truncations were set at min and max values. For probability of delay, a uniform distribution was applied since there is not an apparent "most-likely" value to set as the mean. Probability of LTFU during delay and probability of HIV transmission were both sampled from normal distributions. Discount rates were not modeled in probabilistic uncertainty analysis. One thousand trials were run and incremental cost-effectiveness ratios (ICERs) and net monetary benefits were calculated in analysis.

This study used the commonly accepted methodology from the WHO Choosing Interventions that are Cost-Effective (CHOICE) project which asserts that one to three times (3x) the per capita gross domestic product (GDP) in the country in which the intervention takes place is the most appropriate willingness-to-pay (WTP) threshold for DALYs averted (WHO 2010). This methodology has been used in several recent HIV-related studies in sub-Saharan Africa, including Kenya (Patel et al. 2017). However, there are several challenges to the WHO-CHOICE methodology as a proxy for a WTP threshold for DALYs averted, including that there is little evidence that those thresholds are actually useful to decision-makers and that they may be less relevant for LMIC settings where external donors supplement national budgets but focus on their own agendas (Marseille et al. 2015, Leech et al. 2018). A recent study focusing on opportunity cost estimated that the WTP threshold in Kenya was closer to $73-$1,164 (Woods et al. 2016). To address these concerns, we used the uncertainty analysis trial data to develop a cost-effectiveness acceptability curve for POC at escalating WTP thresholds per DALY averted.

3. RESULTS

3.1 Model Outcomes

Discounted 10-year outcomes for central laboratory and POC VL monitoring approaches for a cohort of 1 million are presented in Table 6. Total costs were higher for implementation of POC VL monitoring ($910,671,705) compared to central laboratory VL monitoring ($873,224,811). POC outcomes were favorable for averting DALYs, deaths, and new transmissions compared with central laboratory VL monitoring. Approximately 68 percent of the cohort for both POC and central laboratory survived after 10 years, with slightly higher survival in the POC approach.

Table 6: Discounted Outcomes over 10-Year Horizon

Horizon	Approach	Total Cost	Total DALYs	Total Deaths	Total New Transmissions
10 Years	Laboratory	$873,224,811	5,023,944	322,216	24,110
	POC	$910,671,705	5,010,337	321,360	23,421

Table 7 presents incremental model outcomes for POC compared to central laboratory. Implementation of POC VL monitoring for one million people over 10 years cost $37.4 million more than the current laboratory approach. However, over 10 years, implementation of POC also resulted in 13,606 DALYs averted, 855 deaths averted, and 690 transmissions averted compared to laboratory VL monitoring.

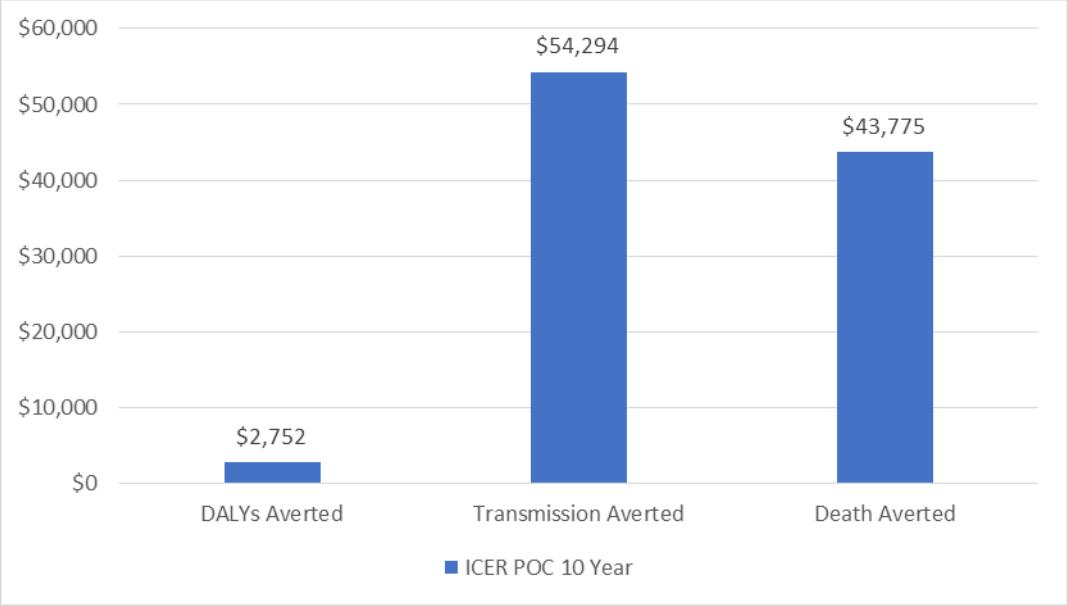
Table 7: Incremental Model Outcomes

Horizon	Approach	Incremental Cost	Incremental DALYs Averted	Incremental Deaths Averted	Incremental Transmissions Averted
10 Years	Laboratory	---	---	---	---
	POC	$37,446,894	13,606	855	690

Central laboratory is the reference approach

The ICERs for POC over a 10-year time horizon are presented in Figure 5 using central laboratory as the reference approach. Using the WHO-CHOICE methodology, the WTP threshold for DALYs averted would be between one to three times the per capita GDP of Kenya. The 2016 per capita GDP in Kenya was $1,455 and three times the GDP is $4,365 (World Bank 2016). Using the high end of the range ($4,365), the cost per DALY averted in the POC approach would be considered cost effective. Using the same WHO-CHOICE methodology, which qualifies interventions that cost below one times (1x) the per capita GDP per DALY averted as "very cost-effective," the POC approach would not qualify as "very cost-effective."

Figure 5: POC Incremental Cost-Effectiveness Ratios over 10 years



Central laboratory is the reference approach

In a secondary analysis, 25,000 individuals newly diagnosed with HIV entered the model through the *Entry FLART* state each cycle reflecting the UNAIDS 2016 estimate that 56,000 adults age 15 and older are infected each year in Kenya (UNAIDS 2016). This approach also captures some of the individuals who may have been LTFU and may re-enter the health system through a new facility. While the first analysis above offers an important estimate for a specific cohort size (1 million) and will be used for all of the following sensitivity and uncertainty analyses, this second analysis may offer a more accurate estimate of the total cost of the program over 10 years, where cohort size fluctuates.

Table 8 provides the results from this secondary analysis. Overall, the total costs for both laboratory and POC were approximately $200 million and $215 million more expensive, respectively, with the addition of new cohort entries each cycle, resulting in total program costs of over 1 billion dollars for both approaches. Incremental cost of POC as well as incremental DALYs averted were higher, as was the ICER per DALY averted ($3,392). However, this secondary ICER was similar to the primary ICER ($2,752) in that it was less than the upper WTP threshold of $4,365, indicating POC is cost effective compared to central laboratory, but higher than the lower WTP threshold of $1,455, indicating that POC was not "very cost effective" according to the WHO definition.

Table 8: Discounted Outcomes over 10-Year Horizon with Added Cohort Entries

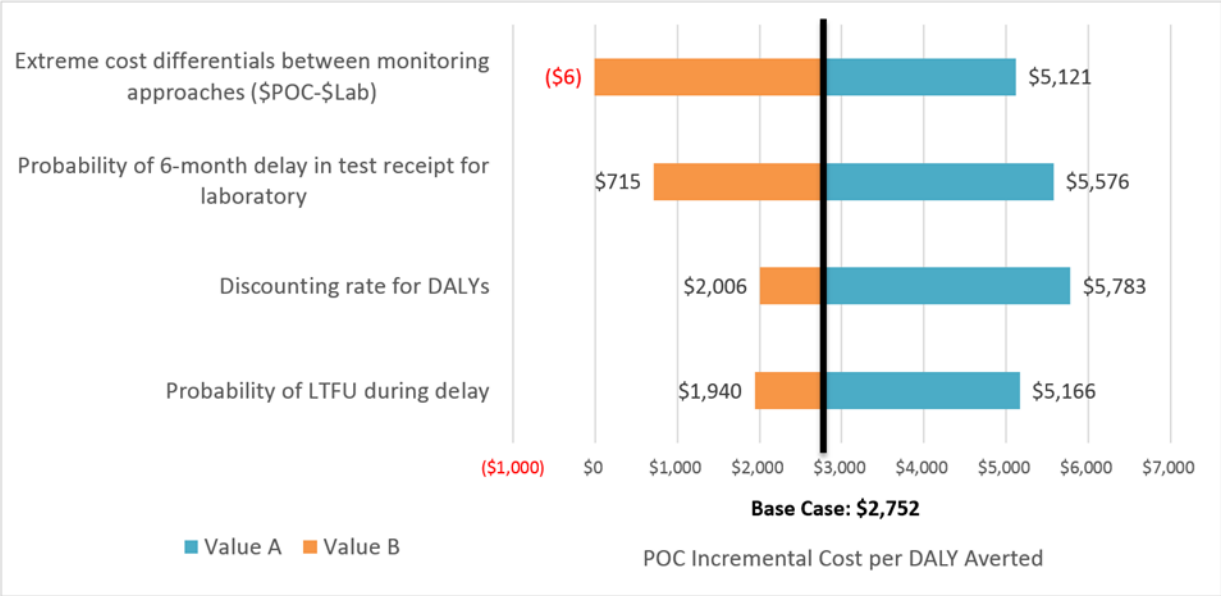
Horizon	Approach	Total Cost	Incremental Cost	Incremental DALYs Averted	Incremental Cost-Effectiveness Ratio per DALY Averted
10 Years	Laboratory	$1,073,317,852	---	---	---
	POC	$1,126,489,418	$53,171,566	15,676	$3,392

Central laboratory is the reference approach

3.2 Deterministic Sensitivity Analysis

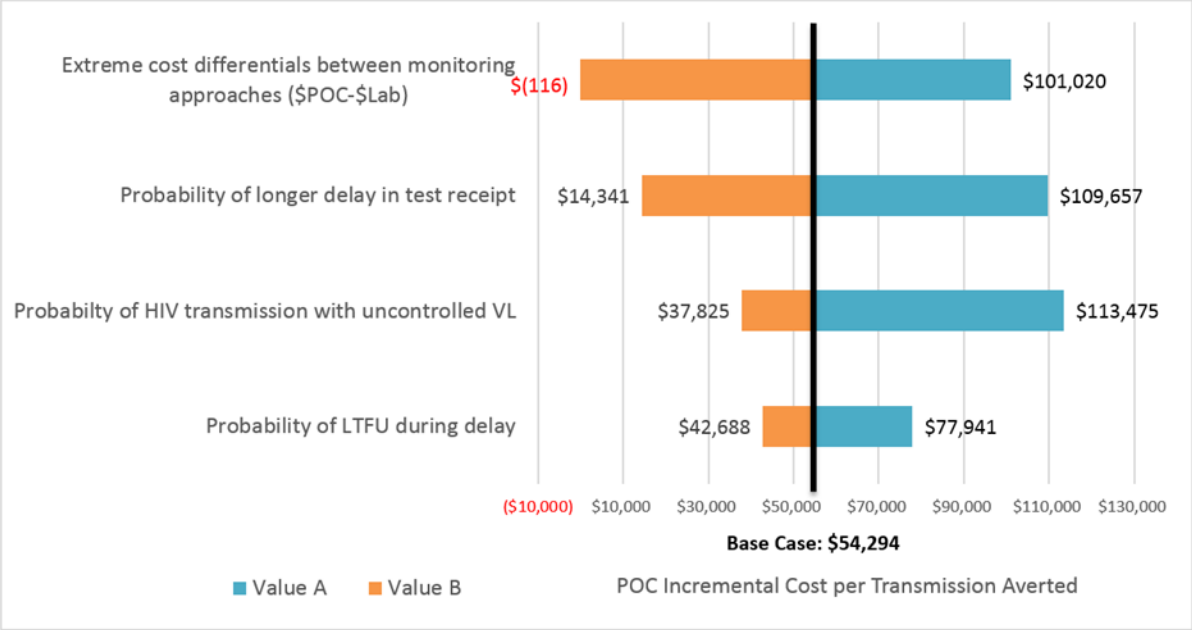
Results of one-way (deterministic) sensitivity analyses are presented in Figure 6. The incremental costs for DALYs averted over 10 years were most sensitive to the cost differential between monitoring approaches, followed by differences in the probability of six-month delay in test receipt. When the lowest per-test cost for POC and the highest per test cost for central laboratory approaches were used (Value B), the POC approach was actually less costly overall, due to the end-of-life costs of the additional deaths in the central laboratory model. This meant that the POC approach was dominant (both less costly and more effective) in that scenario (-$6). When the highest per test cost for POC and the lowest per test cost for central laboratory approaches were used (Value A), the POC approach was not found to be cost effective at $5,121 per DALY averted, using the WTP threshold of $4,365. For all of the Value A parameters (parameters making POC less favorable compared to laboratory), the cost per DALY averted was over $5,000 and would not be considered cost effective. For all of the Value B parameters (parameters making POC more favorable compared to laboratory), POC became more cost effective compared to central laboratory. In the scenario where the probability of six-month delay in test results for laboratory monitoring was high (0.081), the cost per DALY averted was $715, which is "very cost-effective" using the WHO WTP threshold per DALY averted of per capita GDP ($1,455).

Figure 6: Tornado Diagram for POC Cost per DALY Averted at 10 years



As a second set of deterministic sensitivity analyses, we looked at the outcome of cost per HIV transmission averted (Figure 7). For this we also included sensitivity analysis for the probability of transmission due to uncontrolled VL in a delay state, increasing and decreasing the base case parameter (0.010) by 50 percent (0.015 and 0.005, respectively). Similar to the cost per DALY averted sensitivity analysis, the model was most sensitive to changes in the cost differential between POC and laboratory approaches, followed by changes in the probability of six-month delay in test receipt. The model was more sensitive to changes in the probability of HIV transmission than changes in the probability of LTFU during delay. There is not a universal methodology for calculating the WTP threshold for HIV transmissions averted so it is not appropriate to comment on whether these outcomes are cost effective. However, when the cost differences between approaches were minimized, POC was the dominant approach in that it cost less (-$116 per DALY averted) and averted more transmissions, making POC the clear choice.

Figure 7: Tornado Diagram for POC Cost per HIV Transmission Averted

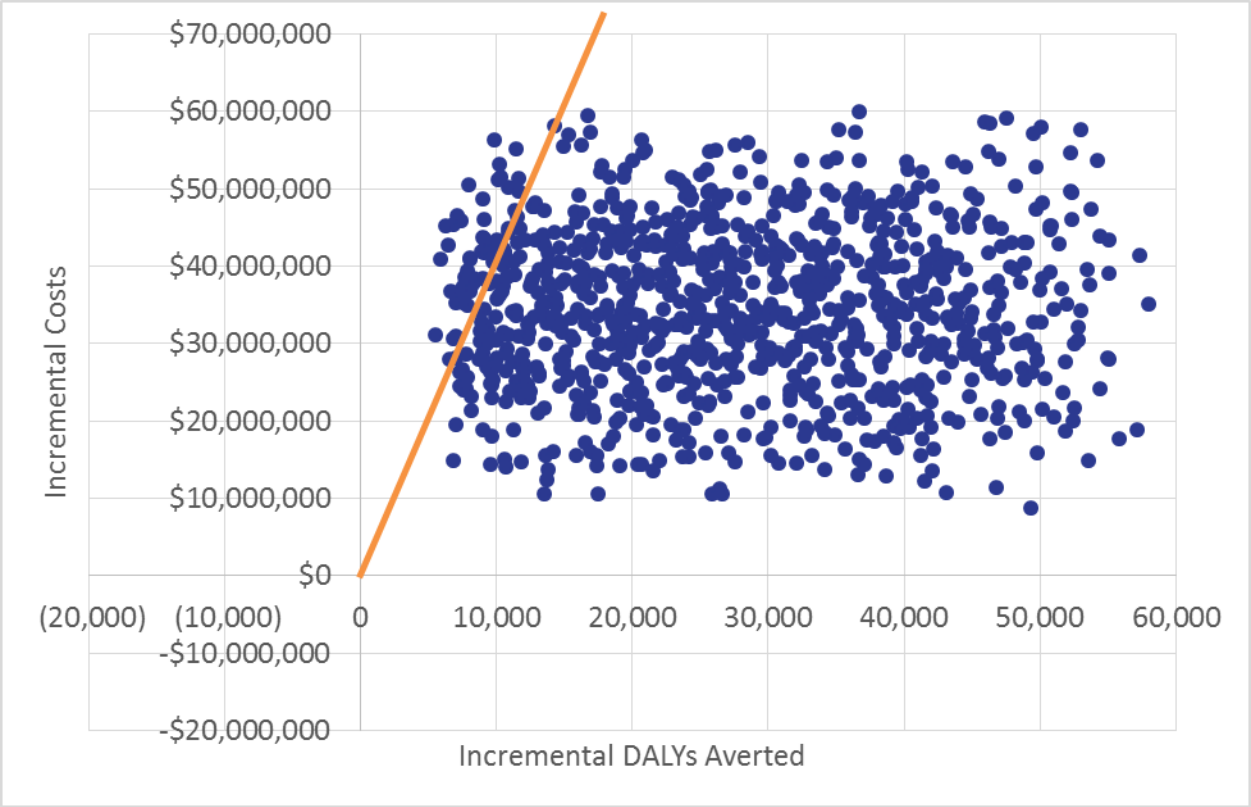


3.3 Probabilistic Uncertainty Analysis

Because the plausible parameter inputs for the one-way sensitivity analyses left some ambiguity as to whether POC would be cost effective in different scenarios, we conducted a probabilistic uncertainty analysis to better assess the likelihood that POC would be cost effective compared to laboratory monitoring given the uncertainty in the model. Figures 8 and 9 present the results from these trials for cost per DALYs averted and cost per transmissions averted, respectively, for POC approach.

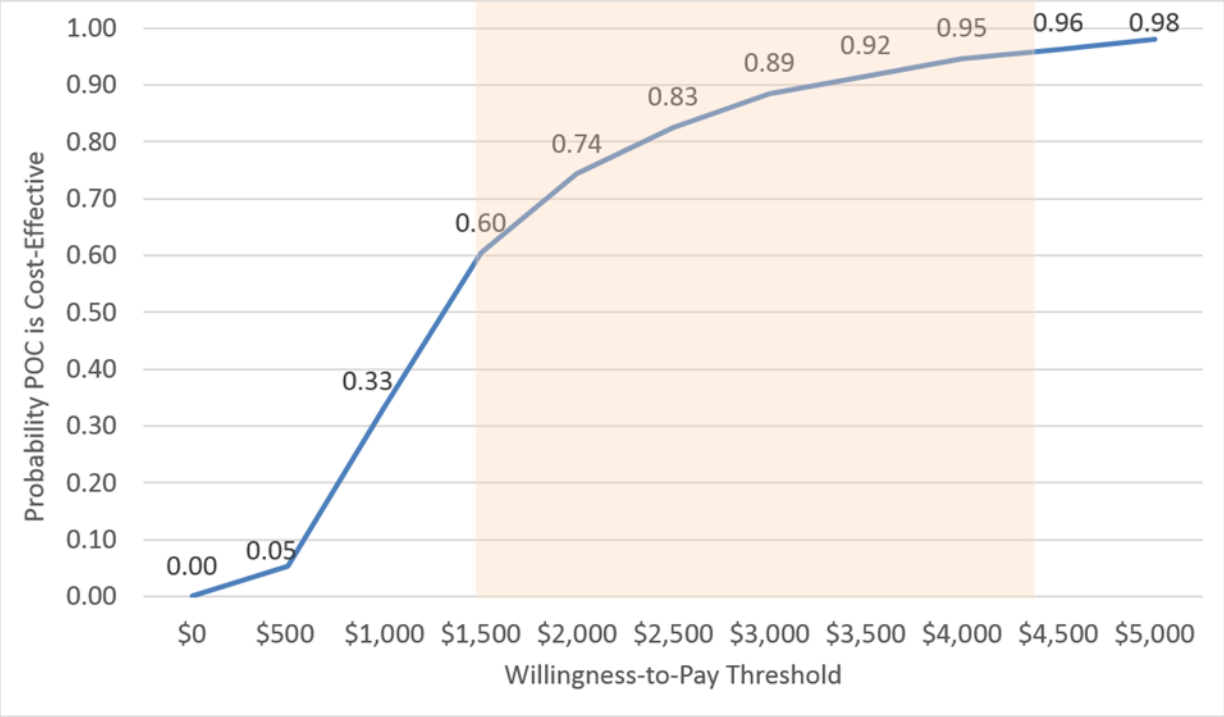
In probabilistic uncertainty analysis, 96 percent of the 1,000 trials resulted in an incremental cost per DALY averted of less than $4,365 (Figure 8), indicating very high probability of cost-effectiveness of POC compared to laboratory VL monitoring over 10 years. Unlike in deterministic sensitivity analysis, no trials indicated that POC was the dominant approach, both averting more DALYs and costing less than central laboratory VL monitoring. The absence of dominant trial values reflects the very small probability of POC and central lab cost parameters simultaneously reaching extreme opposites.

Figure 8: ICER Plane for POC DALYs Averted



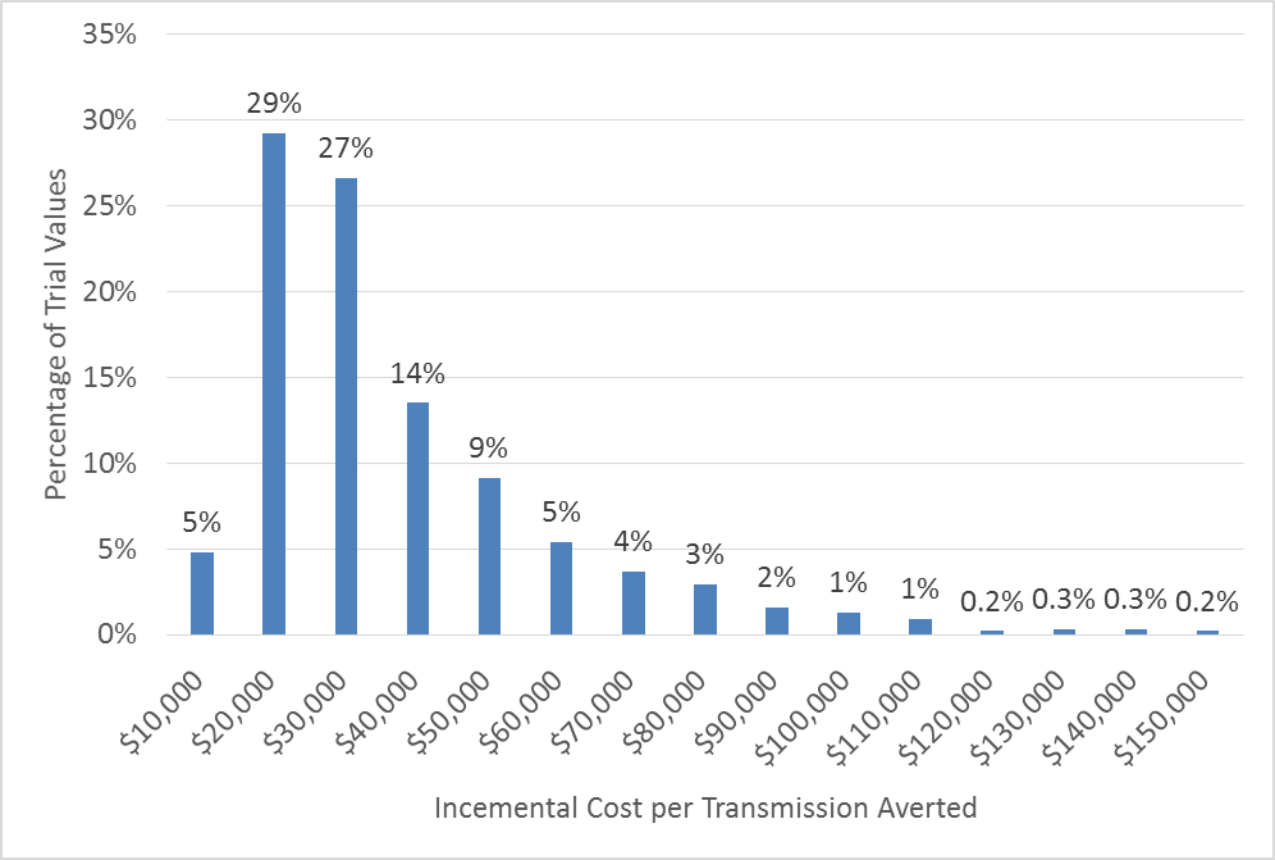
As noted in the Methodology section, there is some debate about what the WTP threshold might be for DALYs in different settings. To address these concerns, we used the uncertainty analysis trial data to develop a cost-effectiveness acceptability curve for POC at escalating WTP thresholds per DALY averted (Figure 9). The cost-effectiveness acceptability curve presents the probability that POC will be cost effective compared to central laboratory given different WTP thresholds (cost-effectiveness thresholds) using cost per DALYs averted as the outcome over 10 years. At a WTP threshold of the 2016 Kenyan per capita GDP ($1,455), the probability that POC would be cost effective compared to central laboratory VL monitoring given the uncertainty in the model is 59 percent. At a two times the per capita GDP ($2,910), it is 88 percent probable that POC would be cost effective compared to central laboratory, and at three times the per capita GDP ($4,365), the probability that POC would be cost effective is 96 percent. As noted earlier, other studies have estimated that the WTP threshold in Kenya is closer to $73–$1,164 which would make the probability that POC is cost effective compared to central laboratory less than 50 percent for the upper value ($1,164) to less than 1 percent for a WTP threshold of $73 (Woods et al. 2016).

Figure 9: POC Cost-Effectiveness Acceptability Curve for DALYs Averted



There is not a universally accepted cost effectiveness threshold for cost per HIV infection averted without calculating the associated DALYs, which this modeling exercise did not do (Jacobsen and Walensky 2017). However, it is still worth considering the incremental cost per transmission averted as a secondary outcome because HIV transmission has a direct impact on the health of the larger Kenyan population and the future demand for HIV monitoring and treatment services. Figure 10 summarizes the results of the 1,000 trial values of ICERs for transmissions averted in POC. Five percent of trial values resulted in ICERs of $0 to $9,999 per transmission averted and 61 percent of trials values resulted in ICERs less than $30,000 per transmission averted. Unlike in the deterministic sensitivity analysis, no ICER values were negative, indicating that POC was never a dominant approach in this uncertainty analysis. Cost-effectiveness of HIV prevention strategies may be better assessed by understanding existing health system priorities and expenditures.

Figure 10: Density Graph of ICERs for Transmissions Averted in POC



4. DISCUSSION

4.1 Key Takeaways

This model compared the costs and outcomes of POC VL monitoring and central laboratory VL monitoring in Kenya, a setting in which central laboratory monitoring has already been implemented at scale. The NASCOP VL monitoring program has made significant progress in the treatment of ART patients and the prevention of HIV transmission, and it has been a flagship for how to support routine monitoring in sub-Saharan Africa. Part of the success of this program has been due to its willingness and ability to evolve with new guidelines and new technologies that best serve the large populations that are HIV positive. POC technology is one of several new innovations that promises to improve the quality, responsiveness, and timeliness of treatment, as well as prevent secondary transmission. This analysis resulted in several important findings, which are presented below.

1. **POC was cost effective compared to central laboratory**

   Using values from the literature and best estimates for model parameters, this model found that POC VL monitoring was cost effective over a 10-year time horizon compared with the current central laboratory VL monitoring approach, at a WTP threshold of three times the 2016 per capita GDP ($4,365). At a cost of $2,752 per DALY averted, POC VL monitoring was well within the WTP threshold, though selecting this approach would still require priority assessment by the Ministry of Health as additional resources are required to implement the POC approach.

   The probabilistic sensitivity analysis supported the base case findings and indicated that the probability that POC was cost effective compared to central laboratory was very high, even given a range of distributional uncertainty. The indifference point of the POC vs. central laboratory approach to VL monitoring was a WTP threshold of $1,250, which is less than the 2016 per capita GDP in Kenya ($1,455) and relatively low by WHO standards. At WTP thresholds higher than $1,250, it was increasingly probable that POC would be the preferred approach.

2. **The model was most sensitive to cost differences in approaches**

   Even though the uncertainty around costs for each approach ranged by only a few dollars and did not overlap, extremes in these cost parameters led to different conclusions in the deterministic sensitivity analysis (ranging from POC dominance to POC not being cost effective). Costs of POC testing are based on applicable primary data, information reported by the Global Fund (2017), and assumptions of POC-platform use under optimal circumstances. The baseline unit cost in this model assumes POC testing is conducted by existing staff at a given health facility and at the maximum daily capacity of the testing platform. Utilization of the POC platforms for non-VL monitoring tests, such as early infant diagnosis of HIV on Alere q or TB detection or STI diagnosis on Cepheid GeneXpert® IV, may lower unit costs for POC VL. Importantly, procurement agreements between countries, manufacturers/suppliers, and third parties may result in different equipment and reagent costs that could impact the cost-effectiveness of POC testing relative to laboratory monitoring. As POC and central laboratory costs shift, it will be important to re-run these analyses to continue to make informed, long-term investment decisions.

3. **The model was also sensitive to probabilities in delay**

   The delay parameter has substantial uncertainty because of differing calculations from NASCOP and HFG primary data collection. In this model, delay leads to higher probabilities of LTFU, which leads to poorer health outcomes, impacting DALYs. Delay is also linked to increased probability of HIV transmission, which does not affect DALYs or costs in this model, but is important to consider from the perspective of the larger health system. With minimal delay in central laboratory test results, POC is not cost effective compared to central laboratory, which otherwise functions very well. However, with larger delays in the central laboratory approach, POC is very cost effective.

   This model assumes that there is no delay in test receipt for POC VL testing. While there is no data available about the probability of delay in the POC approach because it has not been implemented at scale, it is unlikely that POC would be perfectly implemented given the realities of working in often challenging settings. For example, if a machine malfunctions or there is high volume at the clinic, the POC VL monitoring approach may indeed produce delays. To account for this in the model, the probability of delay in the central laboratory approach was minimized to 0.011, but our assumption is that it is unlikely that the POC approach will have more delays relative to central laboratory given the inherent transportation challenges in the central laboratory approach. Pilot data for POC VL monitoring would be useful in assessing the precise difference in test receipt delay.

4.2 Limitations

This model is a simplification of the complex factors that must be considered in the implementation of a VL monitoring program at scale in Kenya. As such, it has several limitations.

One limitation of this model is that co-morbidities and deaths unrelated to HIV were not specifically modeled, and there was no discernment of individual health states (often measured by CD4 counts) within model states. However, these factors would not differentially affect either approach and so they were assumed to cancel out over the large population receiving VL monitoring.

A second limitation of this analysis was that it did not consider the sunk costs already invested in the central laboratory system, the additional costs that would be incurred by switching from the existing central laboratory monitoring approach to POC monitoring, or how central laboratory resources would be reallocated. These costs, associated with training, logistics, new equipment, and new national policies, would likely be substantial in the first five years as a new monitoring approach was brought to scale. However, it is expected that these costs would decline substantially in the long run and given that there are costs associated with any changes to status quo, they were not accounted for in this model.

A third limitation is that costs of central laboratory testing and relevant components used for POC testing are based on a costing study focused in Western Kenya, where the HIV epidemic is most concentrated. While these costs should be largely representative of costs throughout Kenya, given that nationwide procurement of equipment, reagents, and other supplies is all handled by the Kenya Medical Supplies Agency, and human resources for health in public facilities are paid according to national salary scales, variations may still exist between regions of Kenya with differing human resource situations, transportation needs, and levels of HIV prevalence. Another limitation related to cost is that the baseline unit cost in this model assumes POC testing is conducted by existing staff at a given health facility and at the maximum daily capacity of the testing platform. Costs are certain to fluctuate based on the efficiency and usage of both POC and central laboratory machines.

Finally, this model did not consider a hybrid POC-central laboratory approach, which may be the most likely implementation of the new technology, at least initially. In large, established clinics in teaching hospitals in urban settings, central laboratory monitoring may demonstrate superior economies of scale. POC VL testing may be particularly useful for hard-to-reach populations such as injection drug users, sex workers, or rural populations. The relative portability and immediacy of POC systems would support their use in community outreach services and empower clinicians and patients to manage ART appropriately, while minimizing the need for frequent clinical contacts (Dorward et al. 2018). The POC machines could be used to complement the existing established central laboratory approach in places where infrastructure is poor and/or where targeted outreach is warranted. Further, POC technology is not the only innovation in development that can improve the expediency of VL testing and the quality of patient care. Fingerprick and DBS testing are currently not available on Xpert HIV-1 VL and assays with DBS currently have suboptimal performance, but these technologies may offer promising alternatives as technology advances (Pannus et al. 2016).

4.3 Conclusion

This model found that the implementation of POC VL monitoring is cost effective compared to the current approach of central laboratory VL monitoring over a time horizon of 10 years. The key differentiating factors in the two approaches were the costs and the probability of a clinically significant (6-month or longer) delay in patients receiving their test results. While POC VL was the more expensive approach to implement, costing $37 million more over 10 years, it also resulted in fewer DALYs (-13,606), transmissions (-690), and deaths (-855). Despite some uncertainty in the input parameters, the incremental cost per DALY averted ($2,752) was found to be cost effective in most scenarios. These findings highlight the fact that while the current central laboratory VL monitoring program functions well, newer technologies can improve the expediency and effectiveness of VL monitoring, resulting in better patient care and HIV control. The Kenyan Ministry of Health and NASCOP leadership should consider piloting a POC VL monitoring approach, particularly in hard-to-reach settings and populations where transport and continuity of care pose significant obstacles to effective VL monitoring. Pilot data could also be used to make stronger and more precise predictions and recommendations for investments in POC VL monitoring at scale.

It is likely that costs of POC VL monitoring will decrease as more platforms and assays are certified for use and as countries, manufacturers, and international organizations reach procurement agreements similar to those available for lab-based testing platforms. Lower costs, combined with the flexibility afforded by POC technology, in terms of scale and location, and the potential to achieve better outcomes for patients through a more patient-centered model of care should make POC VL monitoring cost effective in settings beyond Kenya. Such developments will aid countries in scaling up or establishing VL monitoring programs tailored to their unique health needs and financial capacities in pursuit of the "90-90-90" goals and an AIDS free generation, even as donor priorities in health transition or funding declines.

REFERENCES

Ajose, O, Mookerjee, S, Mills, EJ, et al. 2012. Treatment outcomes of patients on second-line antiretroviral therapy in resource-limited settings: a systematic review and meta-analysis. *AIDS* 26(8): 929-938.

Arnessen, R, Moll, P, and Shenoi, S. 2017. Predictors of loss to follow-up among patients on ART at a rural hospital in KwaZulu-Natal, South Africa. *PLOS One* 12(5): e0177168.

Ayah, R. 2018. Scaling up implementation of ART: Organizational culture and early mortality of patients initiated on ART in Nairobi, Kenya. *PLOS One* 13(1): e0190344.

Biset Ayalew, M. 2017. Mortality and Its Predictors among HIV Infected Patients Taking Antiretroviral Treatment in Ethiopia: A Systematic Review. *AIDS Research and Treatment*, Volume 2017.

Barnabas, RV, Revill, P, Tan, N, et al. 2017. Cost-effectiveness of routine viral load monitoring in low- and middle-income countries: a systematic review. *Journal of the International AIDS Society* 20(S7): 50-61.

Billioux, A, Nakigozi, G, Newell, K, et al. 2015. Durable Suppression of HIV-1 after Virologic Monitoring-Based Antiretroviral Adherence Counseling in Rakai, Uganda. *PLOS One* 10(5): e0127235.

Boender, TS, Hamers RL, Ondoa P, et al. 2016. Protease inhibitor resistance in the first 3 years of second-line antiretroviral therapy for HIV-1 in sub-Saharan Africa. *The Journal of Infectious Diseases* 214(6): 873-883.

Bonner, K, Mezochow, A, Roberts, T, et al. 2013. Viral load monitoring as a tool to reinforce adherence: a systematic review. *Journal of Acquired Immune Deficiency Syndromes* 64(1): 74-74.

Briggs, A, Sculpher, M, and Claxton, K. 2006. *Decision modelling for health economic evaluation.* Oxford: Oxford University Press.

Brooks, K, Diero, L, DeLong, A, et al. 2016. Treatment failure and drug resistance in HIV-positive patients on tenofovir-based first-line antiretroviral therapy in western Kenya. *Journal of the International AIDS Society* 19(1): 20798.

Brown, LB, Havlir, DV, Ayieko, J, et al. 2016. High levels of retention in care with streamlined care and universal test and treat in east Africa. *AIDS* 30(18): 2855-2864.

Calmy, A, Ford, N, Hirschel, B, et al. 2007. HIV viral load monitoring in resource-limited regions: optional or necessary? *Clinical Infectious Diseases* 44(1): 128-134.

Carter, M. 2017. *A quarter of people with second-line ART treatment failure in Kenya have exhausted current treatment options.* Aidsmap.com. http://www.aidsmap.com/A-quarter-of-people-with-second-line-ART-treatment-failure-in-Kenya-have-exhausted-current-treatment-options/page/3133091/. Accessed April 26, 2018.

Chung, M, Richardson, B, Tapia, K, et al. 2011. A Randomized Controlled Trial Comparing the Effects of Counseling and Alarm Device on HAART Adherence and Virologic Outcomes. *PLOS Medicine* 8(3): e10000422.

CIA World Factbook. 2016. *Central Bank Discount Rate*. https://www.cia.gov/library/publications/the-world-factbook/fields/2207.html. Accessed May 7, 2018.

Cintron, C, Mudhune, V, Haider, R, et al. 2017. *Costs of HIV Viral Load and Early Infant Diagnosis Testing in Kenya*. Bethesda, MD: Health Finance and Governance Project, Abt Associates Inc.

Dorward, J, Drain, P, and Garrett, N. 2018. Point-of-care viral load testing and differentiated HIV care. *The Lancet HIV* 5(1): e8-e9.

Estill, J, Egger, M, Blaser, N, et al. 2013. Cost-effectiveness of point-of-care viral load monitoring of ART in resource-limited settings: Mathematical modelling study. *AIDS* 27(9): 1483-1492.

Ettiègne-Traoré, V, Tuho, M, Mohamed, F, et al. 2013. Differences Between HIV-Infected Men and Women in Antiretroviral Therapy Outcomes — Six African Countries, 2004–2012. *Morbidity and Mortality Weekly Report* 62(47): 946-952.

Fox, M, Bor, J, MacLeod, W, et al. 2016. "Is Retention on ART Underestimated Due to Patient Transfers? Estimating System-Wide Retention Using a National Labs Database in South Africa." Abstract at 21st International AIDS Conference, Durban, South Africa.

Garrett, N, Drain, P, Werner, L, et al. 2016. Diagnostic Accuracy of the Point-of-care Xpert® HIV-1 Viral Load Assay in a South African HIV clinic. *Journal of Acquired Immune Deficiency Syndromes* 72(2): e45.

Geng E, Glidden, D, Bwana M, et al. 2011. Retention in Care and Connection to Care Among HIV-Infected Patients on Antiretroviral Therapy in Africa: Estimation via a Sampling-Based Approach. *PLOS One* 6(7): e21797.

Global Burden of Disease Collaborative Network (GBD). 2013. *Global Burden of Disease Study 2013 (GBD 2013): Disability-Adjusted Life Years and Healthy Life Expectancy 1990-2013*. Seattle, United States: Institute for Health Metrics and Evaluation (IHME).

Global Burden of Disease Collaborative Network (GBD). 2017. *Global Burden of Disease Study 2016 (GBD 2016): Disability-Adjusted Life Years and Healthy Life Expectancy 1990-2016*. Seattle, United States: Institute for Health Metrics and Evaluation (IHME).

Global Fund (The). 2015. *New Approach on HIV Viral Load Testing:* https://www.theglobalfund.org/en/news/2015-06-10-new-approach-on-hiv-viral-load-testing/. Accessed April 26, 2018.

Global Fund (The). 2017. HIV Viral Load and Early Infant Diagnosis Selection and Procurement Information Tool. Version 2. https://www.theglobalfund.org/media/5765/psm_viralloadearlyinfantdiagnosis_content_en.pdf. Accessed April 26, 2018.

Gunda, D, Nkandala, I, Kilonzo, S, et al. 2017. Prevalence and Risk Factors of Mortality among Adult HIV Patients Initiating ART in Rural Setting of HIV Care and Treatment Services in North Western Tanzania: A Retrospective Cohort Study. *Journal of Sexually Transmitted Diseases*.

Gupta, R, Hill, A, Sawyer, A, et al. 2009. Virological monitoring and resistance to first-line highly active antiretroviral therapy in adults infected with HIV-1 treated under WHO guidelines: a systematic review and meta-analysis. *The Lancet Infectious Diseases 9*(7): 409-417.

Hamers, R, Sawyer, AW, Tuohy, M, et al. 2012. Cost-effectiveness of laboratory monitoring for management of HIV treatment in sub-Saharan Africa: a model-based analysis. *AIDS* 26:1663–1672.

Hosseinipour, M, van Oosterhout, J, Weigel, R, et al. 2009. The public health approach to identify antiretroviral therapy failure: high-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy. *AIDS (London, England)* 23(9):1127-1134.

Hyle, EP, Jani, IV, Lehe, J, et al. 2014. The Clinical and Economic Impact of Point-of-Care CD4 Testing in Mozambique and Other Resource-Limited Settings: A Cost-Effectiveness Analysis. *PLOS Medicine* 11(9): e1001725.

Jacobsen, M, and Walensky, R. 2017. Modeling and Cost-Effectiveness in HIV Prevention. *Current HIV/AIDS Reports* 13(1): 64-75.

Jani, I, Meggi, B, Vubil, A, et al. 2016. Evaluation of the whole-blood Alere Q NAT point-of-care RNA assay for HIV-1 viral load monitoring in a primary health care setting in Mozambique. *Journal of Clinical Microbiology* 54(8): 2104-2108.

Jobanputra, K, Parker, L, Azih, C, et al. 2015. Factors Associated with Virological Failure and Suppression after Enhanced Adherence Counselling, in Children, Adolescents and Adults on Antiretroviral Therapy for HIV in Swaziland. *PLOS One* 10(2): e0116144.

Kerschberger, B, Boulle, AM, Kranzer, K, et al. 2015. Superior virologic and treatment outcomes when viral load is measured at 3 months compared to 6 months on antiretroviral therapy. *Journal of the International AIDS Society* 18(1).

Kityo, C, Gibb, D, Gilks, C, et al. 2014. High Level of Viral Suppression and Low Switch Rate to Second-Line Antiretroviral Therapy among HIV-Infected Adult Patients Followed over Five Years: Retrospective Analysis of the DART Trial. *PLOS One* 9(3): e90772.

Kumarasamy, N, Madhavan, V, Venkatesh, K, et al. 2009. High frequency of clinically significant mutations after first-line generic highly active antiretroviral therapy failure: implications for second-line options in resource-limited settings. *Clinical Infectious Diseases* 49(2): 306-309.

Langebeek, N, Gisolf, E, Reiss, P, et al. 2014. Predictors and correlates of adherence to combination antiretroviral therapy (ART) for chronic HIV infection: a meta-analysis. *BMC Medicine* 12(1): 142.

Lecher, S, Ellenberger, D, Kim, A, et al. 2015. Scale-up of HIV Viral Load Monitoring—Seven Sub-Saharan African Countries. *MMWR Morbidity and mortality weekly report* 64(46): 1287-1290.

Leech, A, Kim, D, Cohen, J, et al. 2018. Use and Misuse of Cost-Effectiveness Analysis Thresholds in Low- and Middle-Income Countries: Trends in Cost-per-DALY Studies. *Value in Health*, doi: https://doi.org/10.1016/j.jval.2017.12.016.

Levison, JH, Orrell, C, Losina, E, et al. 2011. Early outcomes and the virological effect of delayed treatment switching to second-line therapy in an antiretroviral roll-out programme in South Africa. *Antiviral therapy* 16(6): 853-861.

Marseille, E, Larson, B, Kazi, D, et al. 2015. Thresholds for the cost–effectiveness of interventions: alternative approaches. *Bulletin of the World Health Organization* 93(2): 118-124.

Mberi, M, Kuonza, L, Dube, N, et al. 2015. Determinants of loss to follow-up in patients on antiretroviral treatment, South Africa, 2004–2012: a cohort study. *BMC Health Services Research* 15(1): 259.

Médecins Sans Frontières (MSF). 2013. *How low can we go? Pricing for HIV viral load testing in low- and middle-income countries.* MSF Undetectable – Volume 3. Geneva: MSF Access Campaign.

Médecins Sans Frontières (MSF). 2014a. *Achieving undetectable: What questions remain in scaling-up HIV virologic treatment monitoring?* MSF Undetectable – Volume 6. Geneva: MSF Access Campaign.

Médecins Sans Frontières (MSF). 2014b. *Getting to Undetectable: Usage of HIV Viral Load Monitoring in Five Countries.* MSF Undetectable – Volume 5. Geneva: MSF Access Campaign.

Médecins Sans Frontières (MSF). 2014c. *Supplementary Material for Issue Brief: Achieving undetectable: What questions remain in scaling-up HIV virologic treatment monitoring?* MSF Undetectable – Volume 6. Geneva: MSF Access Campaign.

Médecins Sans Frontières (MSF). 2016. *Making viral load routine: Successes and challenges in the implementation of routine HIV viral load monitoring.* MSF Undetectable – Volume 8. Geneva: MSF Access Campaign.

Mills EJ. 2011. Life expectancy of persons receiving combination antiretroviral therapy in low-income countries: a cohort analysis from Uganda. *Ann Intern Med.* 155, 4.

Mujugira A et al. 2016. HIV transmission risk persists during the first 6 months of antiretroviral therapy. *J Acquir Immune Defic Synr*, online edition. DOI: 10.1097/QAI.0000000000001019.

Murtagh, M. 2013. Viral load: Current technologies and the pipeline, including point-of-care assays. https://pdfs.semanticscholar.org/presentation/576b/1ef8634a36e22d09cc80ec3e6903d30399ab.pdf. Accessed April 26, 2018.

Mwau, M, Syeunda, CA, Adhiambo, M, et al. 2018. Scale-up of Kenya's national HIV viral load program: Findings and lessons learned. *PLOS One* 13(1): e0190659.

Narainsamy, D, and Mahomed, S. 2017. Delays in switching patients onto second-line antiretroviral treatment at a public hospital in eThekwini, KawZulu-Natal. *Southern African Journal of HIV Medicine* 18(1).

National AIDS and STI Control Programme (NASCOP). 2014. *Guidelines on Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Rapid Advice.* Nairobi: Ministry of Health, National AIDS & STI Control Programme.

National AIDS and STI Control Programme (NASCOP). 2016. *Guidelines on Use of Antiretroviral Drugs for Treating and Preventing HIV infection in Kenya 2016*. Nairobi: Ministry of Health, National AIDS & STI Control Programme.

National AIDS and STI Control Programme (NASCOP). 2018. National AIDS/STD Control Programme Dashboard. http://viralload.nascop.org/. Accessed March 6, 2018.

National Institute for Health and Care Excellence (NICE). 2017. Developing NICE guidelines: the manual.

Pannus, P, Claus, M, Gonzalez, MMP, et al. 2016. Sensitivity and specificity of dried blood spots for HIV-1 viral load quantification. *Medicine (Baltimore)* 95(48): e5475.

Patel, A, Kessler, J, Braithwaite, R, et al. 2017. Economic evaluation of mobile phone text message interventions to improve adherence to HIV therapy in Kenya. *Medicine (Baltimore)* 96(7): e6078.

Petersen, ML, Tran, L, Geng, EH, et al. 2014. Delayed switch of antiretroviral therapy after virologic failure associated with elevated mortality among HIV-infected adults in Africa. *AIDS* 28(14): 2097-2107.

Pujades-Rodriguez M, Balkan S, Arnould L, Brinkhof MA, Calmy A. 2010. Treatment failure and mortality factors in patients receiving second-line HIV therapy in resource-limited countries. *JAMA* 304:303-312.

Reidy, W, Agarwal, M, Lamb, M, et al. 2014. "Loss to Follow-Up: Determining outcomes for adults enrolled in HIV services in Kenya." Abstract at Conference on Retroviruses and Opportunistic Infections, March 3-6, 2014, Boston, Massachusetts.

Roberts, T, Cohn, J, Bonner, K, et al. 2016. Scale-up of routine viral load testing in resource-poor settings: current and future implementaton challenges. *Clinical Infectious Diseases* 62(8): 1043-1048.

Rubaihayo, J, Tumwesigye, N, Konde-Lule, J, et al. 2015. Trends and Predictors of Mortality Among HIV Positive Patients in the Era of Highly Active Antiretroviral Therapy in Uganda. *Infectious Disease Reports* 7(3): 66-73.

Safren, A, Mayer, K, Ou, S-S, et al. 2015. Adherence to Early Antiretroviral Therapy: Results from HPTN 052, A Phase III, Multinational Randomized Trial of ART to Prevent HIV-1 Sexual Transmission in Serodiscordant Couples. *Journal of Acquired Immune Deficiency Syndromes* 69(2): 234-240.

Salazar-Vizcaya, L, Keiser, O, Technau, K, et al. 2014. Viral load versus $CD4^+$ monitoring and 5-year outcomes of antiretroviral therapy in HIV-positive children in Southern Africa: a cohort-based modelling study. *AIDS* 28(16):2451-2460.

Sawe, F, et al. 2013. "Superiority of routine viral load monitoring in rural Kenya: the Kericho Clinic-based ART Diagnostic Evaluation (CLADE) Trial." 20th Conference on Retroviruses and Opportunistic Infections, Atlanta, abstract 151.

Shearer, K, Evans, D, Moyo, F, et al. 2016. Treatment outcomes of over 1000 patients on second-line, protease inhibitor-based antiretroviral therapy from four public-sector HIV treatment facilities across Johannesburg, South Africa. *Tropical Medicine & International Health* 22(2): 221-231.

Sollis, KA, Smit, PW, Fiscus, S, et al. 2014. Systematic review of the performance of HIV viral load technologies on plasma samples. *PLOS One* 9(2): e85869.

Stegman P, Ohadi, E, Buzwani, M, Cogswell, H, Avila C. 2015. *Benchmarking Costs for Non-Clinical Services in Botswana's Public Hospitals.* Bethesda, MD: Health Finance and Governance Project, Abt Associates.

Stevens, W, Gous, N, Ford, N, et al. 2014. Feasibility of HIV point-of-care tests for resource-limited settings: challenges and solutions. *BMC Medicine* 12(1): 173-181.

UNAIDS. 2014a. *90-90-90 An ambitious treatment target to help end the AIDS epidemic.* Geneva: Joint United Nations Programme on HIV/AIDS.

UNAIDS. 2014b. *Landmark HIV diagnostic access program will save $150m and help achieve new global goals on HIV.* http://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2014/september/20140925prviralload. Accessed April 26, 2018.

UNAIDS. 2016. *The need for routine viral load testing: Questions and answers*. http://www.unaids.org/sites/default/files/media_asset/JC2845_en.pdf. Accessed April 26, 2018.

UNITAID. 2015. *HIV/AIDS Diagnostics Technology Landscape (5th Edition).* Geneva: UNITAID.

Vandormael, AM, Boulware, DR, Tanser, FC, et al. 2016. Virologic monitoring can be a cost-effective strategy to diagnose treatment failure on first-line ART. *Journal of acquired immune deficiency syndromes* 71(4):462-466.

Woods, B, Revill, P, Sculpher, M, et al. 2016. Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. *Value in Health* 19(8): 929-935.

World Bank (The). 2016. GDP per capita (current US$): https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=KE. Accessed March 6, 2018.

World Bank (The). 2018. Prevalence of HIV, total (% of population ages 15-49): https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS. Accessed March 6, 2018.

World Health Organization (WHO). 2010. *Choosing interventions that are cost effective (WHO-CHOICE)*. http://www.who.int/choice/en/.

World Health Organization (WHO). 2013. *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach.* Geneva: WHO.

World Health Organization (WHO). 2016. Global Price Reporting Mechanism for HIV, tuberculosis and malaria (database). http://apps.who.int/hiv/amds/price/hdd/. Accessed April 26, 2018.

World Health Organization (WHO). 2017. WHO Prequalification of In Vitro Diagnostics PUBLIC REPORT. Product: Xpert® HIV-1 Viral Load with GeneXpert® Dx, GeneXpert® Infinity-48, GeneXpert® Infinity-48s and GeneXpert® Infinity-80 WHO reference numbers: PQDx 0192-070-00, PQDx 0193-070-00, PQDx 0194-070-00, PQDx 0195-070-00. July 2017, Version 2.0. http://www.who.int.libproxy.lib.unc.edu/diagnostics_laboratory/evaluations/pq-list/hiv-vrl/170720_final_pq_report_pqdx_0192_0193_0194_0195_070-00.pdf?ua=1

Yehia, B, Stephens-Shields, A, Fleishman, J, et al. 2015. The HIV Care Continuum: Changes Over Time in Retention in Care and Viral Suppression. *PLOS One* 10(6): e0129376.

Zwahlen, M, and Egger, M. 2006. *Progression and mortality of untreated HIV-positive individuals living in resource-limited settings: Update of literature review and evidence synthesis.* Geneva: Joint United Nations Programme on HIV/AIDS.

Health Finance
& Governance
Expanding Access. Improving Health.

BOLD THINKERS DRIVING
REAL-WORLD IMPACT