COMPARING OUTCOMES OF DRIED BLOOD SPOT AND PLASMA VIRAL LOAD MONITORING FOR HIV TREATMENT IN RESOURCE-LIMITED SETTINGS USING A MARKOV STATE-TRANSITION MODEL

by
Zachary Elijah Adams

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science.

Baltimore, Maryland

August, 2016

ABSTRACT

Background and Objective: People living with HIV who are receiving antiretroviral therapy need to be monitored to evaluate treatment failure. Gold standard plasma viral load is logistically difficult in many resource-limited settings; dried blood spot viral load testing may be a more accessible alternative. A Markov state-transition model was created in order to better evaluate the clinical consequences of this alternative. Outcomes were compared to those of plasma viral load, CD4 immunologic criteria, and clinical criteria for treatment failure.

Methods: A Markov state-transition model was created with two cohorts of 10,000 sub-Saharan African adults, one ART naïve cohort and one ART experienced cohort. Outcomes of each cohort were simulated over 5 years of follow-up. Outcomes of interest were the number of patients who died or were virologically failing after five years, events of interest were cumulative misclassifications over five years.

Results: Dried blood spot viral load testing was 91% as effective as plasma viral load at averting deaths in the ART naïve cohort and 85% as effective in the ART experienced cohort, compared with clinical symptoms monitoring alone. There were more misclassifications with dried blood spot viral load than with plasma viral load. Both dried blood spot and plasma viral load testing lead to fewer deaths and misclassifications than either clinical criteria for treatment failure alone or immunologic criteria. Estimated programmatic costs for plasma viral load and dried blood spot viral load testing were comparable.

Conclusions: Dried blood spot viral load is a good alternative to plasma viral load (when the latter is unavailable), with comparable clinical consequences and costs. Viral load should continue to be the treatment monitoring mode of choice, as clinical and immunologic criteria are inadequate for timely and correct determination of treatment failure.

Readers: Catherine Sutcliffe, Ph.D. and David Dowdy, MD, Ph.D.

ACKNOWLEDGEMENT

I would like to first and foremost thank my wonderful advisers, David Dowdy and Katie Sutcliffe, who tirelessly worked with my "abbreviated" time constraints and incessant requests for feedback. I hope I didn't drive you both crazy. I would like to thank Homayoon Farzadegan and Frank Curriero, who helped me develop my teaching skills this year and earn a couple bucks doing so. I need to again thank my college adviser, Kirk Miller, who continues to listen to and encourage me to do better, even though he technically has no ongoing responsibility for my education or career. I would like to thank my friends from Hopkins, especially Saul Feinstein, Amrita Rao, Molly Deutsch-Feldman, Diana Higuera, Cassie Ott, Ruth Geller, Ruthie Burrows, and my friends from elsewhere, especially Ryan Akman, Eric Tien, David Kaufman, Ben Tipper, Jack Flowers, Ezra Park, Rob Wilkie, Eric Li, Matt Klimuszka, and Griffin Rhodes. I should also probably thank my family, who have had to deal with me for ~~21~~ 24 years now. I would most of all like to thank my girlfriend, Helen Ryan, who has now endured two theses from me and still decided of her own free will to stick around.

"May science give us the courage to do what we must."

- Trey Parker and Matt Stone

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SUPPLEMENT TABLES AND FIGURES

INTRODUCTION

Global prevalence of HIV infection is estimated by the United Nations to be 36.7 million people, with most of the burden in low-and-middle-income countries (LMIC) (UNAIDS 2016). Highly-active anti-retroviral therapy (ART) coverage is estimated to be 46% (UNAIDS 2016). While gains over the last few years have been great (including a more than doubling in coverage in South and East Africa), there is a long way to go to reach the UNAIDS target of "90-90-90" for the year 2020, "whereby 90% of people living with HIV know their HIV status, 90% of people who know their HIV-positive status are accessing treatment and 90% of people on treatment have suppressed viral loads" (UNAIDS 2014).

Individuals receiving ART need to be monitored to predict treatment progress and evaluate treatment failure (Mellors et al. 1997). Treatment failure is determined through one or more of three criteria: virologic criteria (through viral load, the number of RNA copies of the virus per milliliter of blood), immunologic criteria (through CD4 count, the number of CD4 cells per milliliter of blood; referred to here as "CD4"), or clinical criteria (through clinical progression, marked by an HIV-related illness or other clinical sequela; referred to here as "CSM") (Mellors et al. 1997; World Health Organization 2016). An individual with successful treatment will have low or undetectable viral load, minimal decrease in successive CD4 counts, and no clinical progression (World Health Organization 2016). An individual failing treatment should change to different available line of ART in an attempt to improve viral suppression (Mellors et al. 1997; Mee et al. 2008; Rutherford et al. 2014; World Health Organization 2016).

CD4 is the traditional prognostic marker of HIV (Mellors et al. 1997), though viral load is the widely accepted and WHO-officially recommended treatment failure monitoring method when available (World Health Organization 2016). At worst, it is noninferior to CD4 in reducing mortality through treatment monitoring (Boyer et al. 2013; Jourdain et al. 2013; Mermin et al. 2011; Laurent et al. 2011; A. N. Phillips et al. 2008; Tucker et al. 2014), but it is widely viewed as superior (Chaiwarith et al. 2007; Mee et al. 2008; Ferreyra et al. 2012; Rawizza et al. 2011; Rutherford et al. 2014; Shen et al. 2016; Ingole et al. 2013; Hamers et al. 2012). Viral load monitoring increases life expectancy (Hamers et al. 2012), reduces risk of misdiagnosis of treatment failure (Hamers et al. 2012), reduces risk of drug resistance (Hamers et al. 2012; World Health Organization 2016), and may be cost-saving (Hamers et al. 2012). Clinical symptoms should also be monitored (CSM) as standard practice, though it is not an effective method of determining treatment failure (Rutherford et al. 2014; World Health Organization 2016).

Unfortunately, there have been persistent challenges in implementing viral load monitoring in LMIC (Madec et al. 2013; UNAIDS 2016). In addition to the low uptake of treatment itself (UNAIDS 2016), treatment monitoring has an even lower uptake (Neogi et al. 2011). Currently, plasma viral load (PVL) is the gold standard viral load monitoring method of choice (Fajardo et al. 2014). PVL is widely available in the developed world, but is logistically difficult in high-burden LMIC due to the need for phlebotomy to obtain, freezers to preserve, and vehicles to transport plasma samples (Neogi et al. 2011). In addition to complicated sample handling, PVL necessitates a professional laboratory with specialized equipment and trained staff (Neogi et al.

2011). An emerging alternative to PVL is dried blood spot (DBS) viral load. DBS samples eschew some of the logistical difficulty presented by plasma samples, as they do not need to be professionally obtained, carefully preserved, or tenuously transported (Neogi et al. 2011; Rutstein et al. 2015). Laboratory facilities are still required for DBS analysis, but DBS samples can be transported cheaply through municipal mail services (Lofgren et al. 2009).

The logistical advantages of DBS come at the cost of decreased accuracy (Fajardo et al. 2014). Lower sensitivity may result in fewer virally unsuppressed people being successfully identified and moved from first-line ART to second-line ART, whereas lower specificity may result in more virally suppressed people being misidentified as virally unsuppressed and unnecessarily moved from first-line ART to second-line ART. At a population level, it is certain there will be different costs and outcomes associated with different treatment monitoring types. A Markov state-transition model was created here in order to better evaluate the clinical consequences of using DBS for viral load testing and determine the accuracy threshold needed for DBS to be a desirable alternative viral load test when PVL is not available.

METHODS

Overview

A Markov state-transition model was created with two cohorts of 10,000 sub-Saharan African adults, one ART naïve cohort (initiating first-line ART) and one ART experienced cohort (virologically suppressed and on first-line ART). The outcomes of each cohort were simulated over 5 years of follow-up. The primary aims of this analysis were threefold:

1. To compare the clinical consequences of DBS to PVL in order to evaluate acceptability of DBS where PVL is not available.
2. To compare the clinical consequences of DBS to CSM and CD4 in order to quantify the advantages of viral load testing for evaluating treatment failure.
3. To measure the effects of varying assumptions about the accuracy of DBS viral load testing (relative to PVL) on clinical consequences.

True treatment failure was defined as a true plasma viral load in excess of 1,000 copies/mL (also referred to as virologic failure), as consistent with the current WHO criteria for treatment failure (World Health Organization 2016). Measured treatment failure for viral load tests (DBS or PVL) was defined as two consecutive viral load measurements, 3-6 months apart, in excess of 1,000 copies/mL, consistent with the current WHO criteria for viral load testing (World Health Organization 2016). Measured treatment failure for CSM was defined as a positive clinical diagnosis of an HIV-associated illness. Measured treatment failure for CD4 count was defined as a decline to baseline or persistent CD4 count below 100 cells/mL (World Health Organization 2016). Measured treatment failure is

distinguished from true treatment failure, as the tests (measurements) can misclassify.

The main outcome of interest was the number of patients who died after 5 years of follow-up. Events of interest include the number of patients who were upclassified (unnecessarily switched to second-line ART, despite treatment success, due to a false positive test) and the number of patients who were downclassified (unnecessarily kept on first-line ART, despite treatment failure, due to a false negative test). Also of interest was the number of patients who were virologically failing at the end of follow-up, including both those failing on first-line ART and those who experienced true treatment failure while on second-line ART. Total systemic costs for each cohort after 5 years were also estimated as a crude measure of comparing cost effectiveness among the treatment monitoring methods.

Model Structure

The model followed a simplified Markov state-transition model of HIV treatment, monitoring, and disease progression (Figure 1). A Markov model assumes that individuals are divided into states of some process, with a discrete probability of moving to a different state over the time of observation. Future states are dependent only on the current state, so individuals on second-line therapy cannot return to first-line therapy, and individuals who die cannot move to any other state.

The two parallel model components, treatment monitoring and disease progression, together comprise the 25 total possible states (24 living states and the death state) at each time interval (Figure 2). The treatment monitoring component

follows the outcomes of the tests for treatment failure (DBS, PVL, CSM, or CD4), whereas the disease progression follows the true immunologic state of the individual (true CD4 count). At entry to the cohort, all patients are considered to be alive, virally suppressed, and on first-line ART (or starting first-line ART if they are ART naïve). At exit from the cohort, patients are defined by their regimen (first or second-line), viral load (suppressed or unsuppressed), and vital status (alive or dead).

For simplicity, a minimum number of assumptions were made, and for transparency, assumptions and parameters were chosen for source availability to (and ease of modification by) readers. This flexibility lends itself to easier individualized analysis for decision-making, as well as sensitivity analysis. This model is expected to have less fidelity as a trade-off for simplicity, though this loss is hopefully mitigated by its flexibility.

Virologic testing and treatment monitoring occurred at six and twelve months after entry into the cohort, as well as every twelve months thereafter, for ART naïve patients (up to seven measurements). Virologic testing and treatment monitoring occurred every twelve months after entry into the cohort for ART experienced patients (up to six measurements). The treatment and virologic monitoring model component used time steps of six months over the course of the full five-year projection. In each state, the model first assessed the probability of virologic failure, then applied sensitivity and specificity assumptions for the defined treatment monitoring strategy and assumed that all patients would receive treatment based on the test results. For an individual on first-line ART, a positive

test means switching to second-line ART; a negative test means remaining on first-line ART. Individuals who test positive would require a second test in the same time interval before switching to second-line ART. Tests performed during one time interval were assumed to result in any changes necessary by the next time interval (with a 100% treatment line switch rate).

Updates

The model presented in this manuscript was originally constructed by Mark Moses and David Dowdy, and first presented in Costs and Consequences of Viral Load Monitoring with Dried Blood Spots versus Plasma in Resource-Limited Settings, a report for the World Health Organization HIV Treatment Guidelines Development Committee. This report was published May 2015. This model has been adapted and modified substantially for 2016, following an updated review of the literature (see Supplement for details).

State-transition Parameters

Initial distribution of CD4 states in the cohort of ART naïve patients was ascertained from a cohort study conducted in Kenya, Uganda, and Zambia, chosen for their representativeness of high-burden, resource-limited settings (Institute for Health Metrics and Evaluation 2015). Raw data is available on their web repository (http://ghdx.healthdata.org/ihme_data), and is merged from the Access, Bottlenecks, Costs, and Equity Project 2012 in Kenya, Uganda, and Zambia. Data from Ghana is not included due to the lack of chart extraction data. The study by Maduna in South Africa was considered, but discarded because it included a mixture

of patients on and not on ART at baseline, and a stratified distribution was not reported (Maduna et al. 2015). Initial distribution of CD4 states of the cohort of ART experienced patients was calibrated at entry to the CD4 count distribution of ART naïve patients who were on first-line ART and virologically suppressed after 5 years.

Disease progression was modeled by differential risk of transition and mortality at different CD4+ T-cell count states (> 350, 350-200, 199-100, and < 100 cells/mL) and contingent on viral suppression (assumed from treatment success; suppressed individuals gain CD4, unsuppressed individuals lose CD4). Individual state transition and mortality risk was probabilistically dependent on treatment in the current state. Disease progression was assumed to occur at the same rate in individuals with virologic failure as those who are ART naïve (Mellors et al. 1997; Maduna et al. 2015), though mortality was assumed to be influenced by ART, even among unsuppressed individuals (Maduna et al. 2015). A similar, more recent study in sub-Saharan Africa by Cori et al. was considered as an alternative, though the CD4 grouping categories used in that analysis did not line up with this model (Cori et al. 2015). CD4 transition probabilities among suppressed patients were taken from Gabillard et al. and follow the WHO immunological staging system with an additional granular category at the lower end of the spectrum (below simply "AIDS" or <200 cells/mL) (Gabillard et al. 2013).

Viral suppression probabilities were taken from a trio of studies. First line, suppressed to unsuppressed, was taken from Fox et al. (Fox et al. 2012); second-line, suppressed to unsuppressed, was taken from Ajose et al. (Ajose et al. 2012);

and first line, suppressed to unsuppressed in the first 6 months was taken from Barth et al. (Barth et al. 2010).

Transition probabilities which did not line up with the 6-month-per-state timeline used in this model were transformed assuming an exponential CD4 transition distribution to fit the timeline of this model (Drabick et al. 1992).

Mortality probabilities were taken from Maduna et al. using the cohort on ART (Maduna et al. 2015). Gabillard et al. was considered as an alternative (Gabillard et al. 2013), though Maduna et al. was more recent and used a solely African cohort (versus an African and Asian cohort), which was more appropriate for our target analytic population.

The initial model did not assume treatment side effects, toxicity, or intolerance to be a failure of first-line ART, nor did it assume loss of samples or loss of patients to follow-up. Third-line and fourth-line ART were also assumed to be unavailable. If patients switched to second-line therapy, their costs of care and disease progression continued to be projected, but we therefore assumed their treatment monitoring would end (given that there is no other treatment line to be switched to and they cannot switch back to first-line ART). The model also did not account for disability or quality of life adjustments, opportunistic infections, or HIV transmission. Initial distribution, mortality, virologic failure, and state-transition probabilities are reported in Table 1.

Comparisons

DBS, PVL, CSM, and CD4 count were compared in the baseline analysis. Baseline accuracy measurements for DBS and PVL were informed based

on a review of the literature. For the purposes of comparison in this analysis, PVL is assumed to be the gold standard test with a perfect 100% sensitivity and 100% specificity. Accuracies of treatment monitoring methods are reported in Table 2.

Six commercial DBS assays were also compared, including the Abbott Molecular: Abbott RealTime HIV-1 assays with m2000rt platform ("Abbott"), Biocentric: Generic HIV Charge Virale ("Biocentric"), bioMérieux: NucliSENS EasyQ® HIV-1 v2.0 ("bioMérieux"), Siemens Versant: HIV-1 RNA 1.0 kPCR Assay ("Versant"), and Roche Molecular Systems: COBAS® AmpliPrep/COBAS® TaqMan® HIV-1 Test version 2.0 with the SPEX protocol ("Roche SPEX") and the free virus elution protocol ("Roche FVE"). The DBS accuracy measurements were reported by Vojnov et al. at the Clinton Health Access Initiative with the exception of Roche FVE, which was reported by Carmona and Mahlumba (Vojnov et al. 2014; Carmona and Mahlumba 2014; World Health Organization 2014). All viral load assays were all compared at the 1,000 copies/mL cutoff (World Health Organization 2014). Accuracies of these assays are reported in Table 2.

Cost

The model makes simple assumptions about cost in order to maximize decision-making and minimize error, given informal cohort cost estimates. Individuals on ART are expected to make a clinical visit every three months at a cost of about $20 (four visits per year for $80) (Siapka et al. 2014). This cost is inclusive of CSM, as a clinic visit was assumed to be necessary for patients to pick up their medications. CD4 count tests were not assumed to be part of the clinic visit by default, and were priced at $24 (Hyle et al. 2014), though projected CD4 was

reported as part of the model, regardless of whether testing was incorporated. Annual drug cost is assumed to be $113 per year for a patient on first-line therapy and $321 per year for a patient on second-line therapy (Clinton Health Access Initiative 2015).

PVL and DBS testing are both assumed to cost $22 (Medecins Sans Frontieres 2014), though sample collection for PVL costs an additional $3 for phlebotomy and transport, bringing the total testing cost to $25 for PVL and $22 for DBS (Medecins Sans Frontieres 2014). All prices are converted to 2015 USD, and given the abbreviated time frame for this analysis, no additional discount rate has been applied and medicines are assumed to be generic (Clinton Health Access Initiative 2015). Cost outcomes are reported at the cohort level (with minimum significant figures) and are not broken down individually past the parameter input step. Cost parameters are reported in Table 3.

Sensitivity Analysis

Four sensitivity analyses were performed. A one-way sensitivity analysis was performed to measure the effect of varying assumed DBS viral load sensitivity from 80% to 100% in increments of 5%. Any incident that precluded a treatment failure test (e.g. loss to follow-up or loss of sample) could also be considered a part of the sensitivity assumption. These incidents lead to the same outcome as an insensitive test, they result in the assumption of maintenance on whatever treatment line the patient was on.

A second one-way sensitivity analysis was also performed to measure the effect of varying assumed DBS viral load specificity from 80% to 100% in

increments of 5%. Observed specificities varied greatly in commercial assays (World Health Organization 2014), and this provided a way to visualize the change incrementally.

A third one-way sensitivity analysis was performed to measure the effect of varying assumed line switching due to ART intolerance or side-effects per follow-up period (from 0% to 10% in increments of 2%). This was performed for both DBS and PVL to evaluate possible differences in effects between the two tests.

A two-way sensitivity analysis was performed to measure the effects of varying assumed DBS viral load sensitivity and specificity from 80% to 100% in increments of 5%. This analysis offers fairly granular visualization of changes in both screening accuracy parameters.

Software

This model was constructed in and analysis was performed in Microsoft Excel 2016.

RESULTS

Primary Outcomes

Deaths, sorted by treatment failure test, are shown in Figure 3a, cumulative for the cohorts of 10,000 ART naïve and 10,000 ART experienced patients at the end of 5 years of follow-up. Incremental deaths (compared with CSM), are shown in Figure 3b. DBS and PVL both resulted in dramatically fewer deaths in the cohort compared with CSM. DBS was 91% as effective as PVL at averting deaths in the ART naïve cohort and 85% as effective in the ART experienced cohort, compared with CSM.

Deaths, sorted by commercial DBS viral load assay, are shown in Figure 3c, cumulative for the cohorts of 10,000 ART naïve and 10,000 ART experienced patients at the end of 5 years of follow-up. Some tests appeared to outperform the chosen baseline DBS viral load test with regard to averted deaths, others appeared to perform poorly.

Secondary Outcomes

Cumulative misclassifications, sorted by treatment failure test, are shown in Figure 4a, cumulative for the cohorts of 10,000 ART naïve and 10,000 ART experienced patients at the end of 5 years of follow-up. CSM and CD4 had an extremely high number of downclassifications in both cohorts. DBS viral load had a similar number of upclassifications as CSM in both cohorts. Cumulative number virologically failing at the end of follow-up is shown in Figure 4b.

Cumulative misclassifications, sorted by commercial DBS viral load assay, are shown in Figure 4c, cumulative for the cohorts of 10,000 ART naïve and 10,000 ART

experienced patients at the end of 5 years of follow-up. Two assays had an extremely high number of upclassifications in both cohorts.

Cost

Overall cost and incremental cost compared with CSM, sorted by treatment failure test, are shown in Figure 5a, cumulative for the cohorts of 10,000 ART naïve and 10,000 ART experienced patients at the end of 5 years of follow-up. Overall cost for DBS, PVL, and CD4 were similar among all cohorts.

Incremental cost per death averted, compared with CSM, are shown in Figure 5b. Incremental cost per death averted for CD4 in both cohorts was more than $25,000, whereas PVL and DBS were both about $4,000 for the ART naïve cohort, and $10,000 to $15,000 for the ART experienced cohort.

Sensitivity Analysis

There were an additional 20-30 deaths per 10,000 people among the ART naïve and an additional 5-10 per 10,000 people among the ART experienced over 5 years for each 5% decrease in sensitivity, while holding specificity constant, shown in Table 4a. There were an additional 5-30 deaths per 10,000 people among the ART naïve and an additional 5-15 per 10,000 people over 5 years for each 5% decrease in sensitivity, while holding sensitivity constant, shown in Table 4b.

There were an additional 20-30 downclassifications among the ART naïve and an additional 30-40 per 10,000 people among the ART experienced over 5 years for each 5% decrease in sensitivity, while holding specificity constant, shown in Table 4c. There were an additional few hundred upclassifications per 10,000 people

among both the ART naïve and ART experienced over 5 years for each 5% decrease in sensitivity below 95%, while holding specificity constant, shown in Table 4d. Upclassification rate was unchanged by changes in sensitivity, downclassification rate was mostly unchanged by changes in specificity.

There were an additional 30 deaths per 10,000 people among the ART naïve and an additional 20 per 10,000 people among the ART experienced over 5 years for each 2% increase in switch rate per follow-up period due to ART intolerance or side-effects, shown in Figures 7a and 7b.

Changes in deaths, varying both sensitivity and specificity, are shown in Table 4a and 4b for the ART naïve and ART experienced respectively. Changes in upclassifications, varying both sensitivity and specificity, are shown in Table 4e and 4f for the ART naïve and ART experienced respectively.

Changes in downclassifications, varying both sensitivity and specificity, are shown in Table 4c and 4d for the ART naïve and ART experienced respectively. There were slightly fewer downclassifications as specificity decreased.

Changes in programmatic cost, varying both sensitivity and specificity, are shown in Table 4g and 4h for the ART naïve and ART experienced respectively. Crude cost estimates were more affected by changes in specificity than sensitivity. Each 5% decrease in sensitivity led to a maximum increase of $100,000 in programmatic cost, whereas each 5% decrease in specificity led to an increase of $100,000 to $300,000 in programmatic cost for the ART experienced. On the other hand, for the ART naïve, the same effect of specificity was seen, but costs decreased slightly as sensitivity decreased.

DISCUSSION

Clinical Outcomes

The simulation performed here reiterates that PVL and DBS are superior across all outcomes, though add additional cost, considering CSM may be a part of the standard visit to pick up medication refills already. For the additional cost, they provide a lot of life-saving potential (via timely switching to second-line therapy). This is especially true among the ART naive, where the marginal cost per death averted is fairly low.

Furthermore, DBS appears to be an acceptable alternative to PVL, costing only slightly more and resulting in only slightly poorer clinical outcomes. Though it results in some misclassification (something PVL was assumed in this model to not do), the outcomes are drastically better than either CSM or CD4. Treatment experience should be considered when deciding to implement viral load monitoring at the community level. ART experienced groups have better outcomes in general, and are predicted to have a smaller margin for improvement from added virologic testing than the ART naïve. Deleterious primary and secondary outcomes were less frequent and costs were lower overall among the ART experienced cohort. In a setting where a larger proportion of the population is already enrolled in ART follow-up, it may be less important for those individuals to monitor viral loads. There is still a case to be made that if the ART naïve patients are being put on viral load monitoring, ART experienced should be as well (because it is therefore available and already known to be outcome-optimal).

An additional concern at the community level is viral non-suppression. Both upclassification and downclassification lead to greater risk of complete virologic failure, since patients on second-line therapy have 4 times the probability of virologic failure compared with those on first line therapy (Fox et al. 2012; Ajose et al. 2012). This model does not account for transmission, which is mostly prevented when viral loads are suppressed, but is a major community risk from unsuppressed individuals, who transmit most when they have high viral load (Wood et al. 2014; Hamers et al. 2012). This notable exemption from the model leads to a conservative estimate of the importance of viral load monitoring.

For all of these reasons, in addition to decreased risk of ART resistance (something not evaluated in this model), the WHO recommends viral load monitoring for all people on ART, and DBS makes it widely possible (World Health Organization 2016).

Cost

This analysis does not constitute a rigorous, formal cost-effectiveness analysis. There is no incremental cost-effectiveness ratio calculated, nor cost adjustments made for loss to follow-up, quality of life, opportunistic infections, ART regimen adherence, or any other personalized factors. These aggregated cost estimates are not meant be highly precise, or to apply directly to a clinic or country, they are meant to serve as a generally informative tool (with numbers realistically "in the same ballpark").

Cost projections from the model do support a more rigorous cost-effectiveness analysis which found that viral load monitoring was likely to be cost

effective in resource-limited settings compared with CSM and CD4 monitoring (A. Phillips et al. 2015). Phillips et al. did not compare DBS to PVL, the current gold standard (though it projected costs for the latter), but they found that viral-load-informed differentiated care lead to a greater switch rate (i.e. people were more able to switch lines of treatment when a line failed to virologically suppress), which subsequently lead to better health outcomes, that this type of differentiated care was cost effective, and that it was even cost saving when visit frequency was reduced for patients who were virally suppressed (A. Phillips et al. 2015). Cost-savings makes sense when comparing the lower cost of testing to the higher cost of clinical visits. Others have expressed a similar encouragement for reducing the frequency of CD4 measurement for virally suppressed individuals (Hyle, Sax, and Walensky 2013; Girard et al. 2013; Gale et al. 2013; Shen et al. 2016).

CSM was still found to be substantially cheaper than either DBS or PVL for treatment monitoring, but it produced far worse clinical outcomes. Each option needs to be weighed on a contextual basis. Much of this cost gap can be attributed to the higher price of second-line therapy, and individuals on CSM have a lower probability of being indicated by CSM to switch. Most importantly, this research supports the assertion that DBS viral load monitoring is only marginally more expensive than PVL when including the cost of PVL transport and phlebotomy (Medecins Sans Frontieres 2016; A. Phillips et al. 2015).

Sensitivity Analysis

Underestimation of sensitivity assumed can be used as a crude stand-in for other processes in the test-and-treat system, including failure to follow-up, delays in results, and sample loss in transportation or storage. Lower sensitivity, would result in the same outcome as a negative test: no change in treatment line, regardless of whether it is a downclassification. But test downclassifications are a rarer outcome than upclassifications, because a downclassification requires two consecutive false negatives (a very low probability). As specificity decreases, it becomes increasingly easier for an individual to test false positive twice consecutively. This is concerning, especially for DBS viral load testing, where false positives at very low viral loads may be more likely (Monleau et al. 2010; Bertagnolio et al. 2010; Garrido et al. 2009; Arredondo et al. 2012). Increasing the viral load cutoff for a positive test to 5000 copies/mL, rather than the 1000 copies/mL as recommended by the WHO and used in this study, might alleviate this concern (Balinda et al. 2016; Fajardo et al. 2014).

Given that intolerance did not impact PVL and DBS cohorts differentially, it seems as if it is not a particularly important phenomenon to model for comparison. It is likely much more important for cost projection.

Misclassification

Readers and clinicians need to have a discussion about what type and levels of misclassification (among other deleterious outcomes) are tolerable to them. Lower sensitivity results in greater numbers of downclassifications and lower specificity results in greater numbers of upclassifications. This analysis might

show that specificity is more important for two reasons. First, because the downclassifications happen at a much lower rate than upclassifications and vary more readily with changes in their key parameter. Second, because the costs of upclassification are much greater at the full cohort level. This is a comparison from a public health level, and does not evaluate what the differential impact is on death (nor the value of death). Upclassifications impact society more and downclassifications impact individuals more. Even so, downclassifications can be corrected. A downclassification can be uncovered in future treatment monitoring and the individual can have their ART line switched. But an upclassification is permanent, because individuals cannot switch back on ART lines due to the likely emergence of resistance.

This discussion will also play into perceptions about the model used here. Given the field of options of commercial DBS assays, the baseline sensitivity used here may be underestimated, but the baseline specificity used here may be overestimated. If upclassification is a more serious issue, the model may be risky. If upclassification is a less serious issue, the model may be overly conservative.

Dried Blood Spots

There are some logistical challenges of DBS. Heat and humidity can put test reliability at risk (Monleau et al. 2010), and it requires a lab to do the actual analysis, even if a layperson can both draw and transport the sample. But the advantages far outweigh these minor challenges.

Viral load testing provides strong improvement in outcomes over CSM, with CD4 being an insufficient stand-in. Four of the commercial DBS tests appear to be

sufficiently close in outcome to PVL as well: Abbott, Versant, bioMérieux, and Roche FVE. If any of the DBS tests can get close to the 90/90 cutoff assumed here, it would be extremely favorable for cost/benefit ratio (i.e. cost per death averted) and reduction in adverse outcomes compared with CSM, especially in ART naïve cohorts, where the effect is greatest. Good virologic monitoring is a front loading of cost, because the benefits of correct treatment classification reduce deleterious outcomes and may reduce individual costs, considering that the ART naïve cohorts become ART experienced cohorts, and net cost per person decreases as net outcomes improve.

This conclusion echoes that of other recent studies (Templer et al. 2016; Balinda et al. 2016; Rutstein et al. 2015) and a recent systematic review (Smit et al. 2014), which all found strong similarity between DBS viral load and PVL accuracy. These other studies did not simulate, summarize, or compare the costs and outcomes of as many alternative monitoring strategies in one place, but they represent a body of recent literature that supports the use of DBS as necessary (World Health Organization 2014).

Strengths

This model was designed with transparency in mind. The parameters were fairly simple and assumptions fairly limited. This model can and should be modified to fit the needs of the decision-making settings.

What is best for each clinic can and should be evaluated on a case-by-case (setting-by-setting) basis. In places where there is access for PVL, PVL should be used if possible. In places where there is no access for PVL, DBS is likely good

enough. Clinics do not have to choose one or the other, they can mix. The key is that viral load treatment monitoring is not one-size-fits-all, and clinics should make decisions based on what is affordable and feasible, and what is best for their patients. This model is an attempt to project cost and clinical outcomes so that these decisions can be weighed with more information.

Limitations

This model makes an assumption that the underlying processes of HIV viral load testing and disease progression follow the Markov property, that disease states are independent. This may or may not be true in reality. Sensitivity and specificity may be dependent on viral load, rather than being the same across all cases (A. Phillips et al. 2015). Phillips et al. were not alone in this supposition, and the dependency, if real, is most exaggerated at the breakpoint of upclassification/downclassification (Fox et al. 2012; Laprise et al. 2013; Garrido et al. 2009; Monleau et al. 2010; Templer et al. 2016; Arredondo et al. 2012; Balinda et al. 2016). Furthermore, if true, this violates the Markov property, though the degree to which this model as a whole would be rendered inaccurate is unknown.

This model also needs to be modified in order for the results to apply most appropriately to a given context. If one clinic only has access to certain tests and medications with different costs and sensitivity/specificity parameters, they cannot follow the base case analysis here, they need to redo the analysis with modified parameters. And despite the strides toward "simplicity", there is still the concern that the model is not understandable enough. The audience should not just be health researchers with formal statistical training, but health practitioners as well,

who may lack such training. The model has not been deployed to any sorts of field tests or focus groups so there is no real answer to that question.

There are also real concerns about the validity of some of the parameters. Switch rate, which is assumed here to be 100%, is not going to be perfect in practice. This means that regardless of the amount of cases correctly identified as treatment failure, some people will in effect be downclassified by not changing treatment lines. The DBS viral load accuracy parameters, gained from systematic reviews by Vojnov, Carmona and Mahlumba, are unpublished and are cited only by the World Health Organization (World Health Organization 2016).

Additionally, the available treatments and guidelines for treatment would affect the underlying distribution of HIV-infected individuals and their transition properties, which could leave the parameters used here outdated. Over time, if UNAIDS-predicted trends continue (UNAIDS 2016), and more people adopt WHO treatment guidelines (World Health Organization 2016), there will be fewer and fewer ART naïve individuals and the gains of implementing DBS will be fewer.

Finally, and most of all, the cohorts studied here are simulated, not "real". They are extrapolated and built upon by real data, but they are not subject to unpredictability of real people or strength of a prospective clinical or cohort trial. This simulation is predetermined, and does not capture that type of variability.

Further Development

Widening the scope of applicability is the natural starting point for advancing HIV treatment. Viral load monitoring systems need to expand to capture more of the population, clinical and immunologic monitoring are insufficient. Viral load-

informed differentiated care, as modeled here, can save money and lives.If we are to ever meet the 90-90-90 goal ("90-90-90: An Ambitious Treatment Target to Help End the AIDS Epidemic" 2014), we will need better assays with wider availability. As far as this type of model, a multiple cohort comparison of different treatment monitoring methods would go a long way toward validating (or contrasting) the results found in this analysis.

CONCLUSION

The Markov state-transition model used here, which followed simulated cohorts of ART naïve and experienced sub-Saharan African adults, found DBS viral load monitoring to attain similar clinical outcomes and costs to PVL. These results reinforce that DBS is an appropriate viral load monitoring method where PVL is not available, even when assuming conservative accuracy figures. Clinical symptoms monitoring costs less and is easier to perform, but results in poorer health outcomes for patients, including inappropriate switching to more expensive second-line therapy, failure to switch when first-line therapy has failed, reduced likelihood of viral suppression, and increased risk of death. Further research is necessary to validate some of the assumptions made by the model, but outcomes appear to be consistent with other recent research, and the structure and is easier to understand, use, and modify.

SUPPLEMENT

Supplement 1. Overview of the Current State of Treatment Monitoring

The World Health Organization recently released their "Consolidated Guidelines on The Use of Antiretroviral Drugs for Treating and Preventing HIV Infection" (2nd edition), which recommends for the first time (for the WHO) that all people living HIV should receive treatment (antiretroviral therapy) (World Health Organization 2016). Routine viral load testing has been recommended as the ideal for diagnosing treatment failure (World Health Organization 2014; World Health Organization 2016). Viral load monitoring is newer than CD4 monitoring, now aimed at replacing it and supplementing CSM for determining treatment failure (World Health Organization 2016), but it has faces major logistic challenges, especially with requirements for plasma viral load (Neogi et al. 2011; Madec et al. 2013; Fajardo et al. 2014).

Ten studies (Laurent et al. and Boyer et al. followed the same study) (Laurent et al. 2011; Boyer et al. 2013) on the mortality under different methods or tests for virologic failure were included in Supplement S1, spanning publication dates from 2008 to 2015. At minimum, some prior studies found viral load monitoring methods were non-inferior to CSM and CD4 for survival (Tucker et al. 2014; Jourdain et al. 2013; Mermin et al. 2011; A. N. Phillips et al. 2008). Most concluded that they offered superior survival rates (Laurent et al. 2011; Ingole et al. 2013; Shen et al. 2016; A. Phillips et al. 2015; Chaiwarith et al. 2007; Mee et al. 2008). In addition to inferior survival, CSM alone and CD4 count measurement are not

sufficient for determining treatment failure (Ferreyra et al. 2012; Rutherford et al. 2014; Rawizza et al. 2011).

Inability to (quickly) determine treatment failure leads to extended time with unsuppressed viral load (Calmy et al. 2007). This is important because cumulative time with unsuppressed viral load (also known as "viremia copy-years") is positively associated with morbidity and mortality (Cole et al. 2010; Mugavero et al. 2011). Both CSM and CD4 are likely to eventually detect treatment failure, as unsuppressed viral load would lead to deteriorating health and HIV-related sequela, but it would be beneficial to determine failure before these individual outcomes emerge.

There are also broader concerns of transmission and resistance. Individuals with unsuppressed viral replication will transmit HIV at a greater rate, posing a risk to the community (Hall et al. 2013; Calmy et al. 2007). Furthermore, unsuppressed viral load leads to the emergence of drug resistance (Gupta et al. 2009; Bangsberg 2008; Calmy et al. 2007). Drug resistance of an individual's HIV infection jeopardizes their future treatment options, especially in resource-limited settings where there may be fewer lines of treatment available. Drug resistance of an individual's HIV infection may also lead to spread of community drug resistance if they transmit to others. Viral load monitoring provides a way to combat these risks (Calmy et al. 2007; Hall et al. 2013; Gupta et al. 2009; Bangsberg 2008).

Most of all, viral load monitoring measures most directly the results (and success, or lack thereof) of treatment. The goal of treatment is to suppress viral replication. Clinical and immunologic monitoring may detect the result changes in

viral load, but they do not measure viral load. If we want to know if treatment is working, conceptually, the best way to evaluate that is measuring whether the virus is successfully reproducing at a high rate.

Prior simulations found viral load to compare favorably to both CSM alone and CD4 for treatment failure. In 2008, Phillips et al. found viral load monitoring to lead to the greatest survival, though not cost savings (A. N. Phillips et al. 2008). In 2012, Hamers et al. found viral load monitoring to lead to the greatest survival and cost savings, when used at 6 month intervals (compared to CD4 count at 6 month intervals, though not at 12) (Hamers et al. 2012). Most recently, in 2015, Phillips et al. found viral load (specifically DBS) monitoring to lead to the lowest death rate and incidence of HIV, as well as improvement in disability adjusted life years with higher switch rates (A. Phillips et al. 2015). Phillips et al. additionally found likely cost effectiveness with the use of viral-load-informed differentiated ART care with DBS viral load testing (A. Phillips et al. 2015).

The analysis provided in the main write-up primarily aims to evaluate how effectively DBS can be used in place of PVL and follow-up the prior simulation studies above with an updated and refined model for 2016.

Supplement 2. Parameter Revisions

In the process of constructing this analysis, a literature search was performed to update parameters in the model as necessary. The model presented in this manuscript was originally constructed by Mark Moses and David Dowdy, and first presented in May 2015 as "Costs and Consequences of Viral Load Monitoring with Dried Blood Spots versus Plasma in Resource-Limited Settings", a report for the World Health Organization HIV Treatment Guidelines Development Committee. Though the searches were neither automated, nor exhaustive, this model has been adapted and modified substantially for 2016, following an updated review of the literature.

This review was intended to update or verify values for the following parameters:

- CSM, CD4, and DBS treatment failure accuracy
- Side effects rate
- Viral suppression rate (first line, second line, and first line in the first 6 months)
- State transition by CD4 states (suppressed upwards, unsuppressed downwards)
- Mortality by CD4 states
- Initial distribution of CD4 states (ART naive)
- Cost (first line ART, second line ART, PVL, DBS, CSM/clinic)

Supplement Review S1 pertains to mortality parameters for patients in CSM, CD4, DBS, and/or PVL treatment monitoring. Supplement Review S2 pertains to testing accuracy parameters, and summarizes selected studies which reported sensitivity and specificity for CSM, CD4, and/or DBS treatment failure assessments. Cost, initial distributions, side effect rates, and state transitions were not listed in either table due to insufficient sources for comparison. This does not mean that their values remained unchanged from the prior model version.

Inclusion Criteria

Studies to inform the performance of each mode of treatment monitoring were required to report accuracy parameters (sensitivity or specificity values), location (South America, Africa, and Asia only, to maximize representativeness for resource-limited settings), recency (published in the past 10 years, enrolling patients in the past 15 years), mortality rates, and/or threshold of detection (1,000-5,000 copies/mL for viral load threshold only, if reported, as lower values resulted in unreasonably low sensitivity).

Literature Review Search Process

The databases searched were PubMed and Google Scholar. There were no specific journals selected for or searched. Searches took place between May and July 2016.

Keyword search terms:

- Dried Blood Spots | DBS
- Plasma | PVL
- Clinical Symptoms
- Sensitivity | Specificity | Accuracy
- Model(ing)
- Assay
- Test
- Cost
- Outcomes
- Viral Load (Monitoring)
- CD4 (Count) (Monitoring) (Evolution)
- Treatment (Failure) (Monitoring)
- Virologic (Failure) | VF
- (Sub-Saharan) Africa | Resource-Limited

Keywords in parentheses "( )", were used in conjunction with keywords in their row (e.g. 'Africa' and 'Sub-Saharan Africa' were searched terms, but not 'sub-

Saharan'). Numerous parenthetical keywords were often used together (e.g., 'CD4 count monitoring' or 'CD4 count evolution'). Keywords after the pipe, "|", were used as alternates for keywords in their row (e.g. 'Dried Blood Spots' and 'DBS' were searched terms, but never together). Joining multiple piped and parenthetical terms generated complete (and often broad) keyword search strings like 'Dried Blood Spots Viral Load Treatment Monitoring Sub-Saharan Africa'.

Relevant literature was cataloged in the Paperpile reference manager on Google Chrome. Basic information on this literature was transferred to a spreadsheet in Microsoft Excel 2016 for quick access to reference summaries. The most relevant papers were then expanded into tables in this supplement for ease of reading and evaluation. Summary information includes first author, journal, year of publication, type, location, size, outcomes (from results), and conclusion(s) of the studies. Not all references cited in the complete analysis are included in these tables.

Many of these study results were not incorporated directly into the model, but were used to fulfill the literature review's second purpose in informing the hypothesis that DBS was a sufficient substitute for PVL and that both were superior to CD4 and CSM (*for health outcomes). Some of this evidence is presented in S2, based on comparative accuracy measurements, and some is presented in S1, based on comparative mortality outcomes.

Summary and Model Updates

Eight studies on the accuracy of different methods or tests for virologic failure were included in Supplement S2, spanning publication dates from 2007 to 2015. The primary finding of these studies was that CD4 immunologic criteria for virologic failure were as good or better than clinical symptoms monitoring criteria for virologic failure, with accuracy of clinical symptoms monitoring in the 10-15% sensitivity, 80-100% specificity range, and CD4 count monitoring in the 25-35% sensitivity, 80-100% specificity range. Neither the immunologic nor the clinical symptoms monitoring failure assessment performed better than virologic testing in predicting virologic failure and determining treatment failure. Additionally, dried blood spot viral load testing produced similar accuracy results to plasma viral load testing in most scenarios in the 85-100% sensitivity, 85-100% specificity range (Napierala Mavedzenge et al. 2015; World Health Organization 2016; A. Phillips et al. 2015) (with notable exceptions being the low specificity of two of the commercial assays found in Vojnov et al.'s systematic review) (World Health Organization 2016). The previous iteration of this simulation used a 95% sensitivity and 90% specificity, but the current iteration uses a 90% sensitivity and 90% specificity as baseline DBS accuracy values. This is probably an underestimation of sensitivity and a slight overestimation of specificity, given the commercial values from Vojnov et al. (World Health Organization 2016), but are close, round, and easy to modify in sensitivity analysis.

Values for some other parameters, including state transition probabilities remained unchanged due to lack of data that fit the structure of the model (primarily due to the state categories not matching).

For cost updates, the Clinton Health Access Initiative's "ARV Market Report", dated November 2015 and cited by the WHO and others, was used to determine first-line and second-line ART pricing (Clinton Health Access Initiative 2015). The CHAI report provided an "average" cost for both lines in resource-limited generic-accessible countries of $113 for first-line and $321 for second-line therapies (both per person year) (Clinton Health Access Initiative 2015). The most up-to-date pricing available was from Medecin Sans Frontiere's "Untangling the Web of ARV Price Reductions", whose newest edition was dated July 2016 (Medecins Sans Frontieres 2016). The MSF report provided a "lowest" cost and specified a regimen-by-regimen cost, with a first-line of TDF/FTC/EFV priced at $100 and a second-line of AZT/3TC/ATV/r priced at $286 (both per person year) (Medecins Sans Frontieres 2016). The choice to compromise for a set average or estimated value for the simulation was made to allow for the most flexibility and is comparable to the choice to compromise for 90% sensitivity and 90% specificity on DBS testing parameters. Using pricing MSF's cheapest lines of choice is comparable to using the most accurate commercial DBS viral load assay, and it might also show favoritism toward a certain line (or assay). It is less important that the numbers are perfectly accurate and more important that they are conservative and more broadly usable (especially for pricing).

Supplement Tables T1, T2, and T3 contain final model parameters and sources, with updated numbers marked and short notes on the updates below. Some parameter values were simply recalculated due to error or modification of parameters they were dependent on. Some sources were changed to reflect the primary source rather than a secondary source, even if the value(s) remained unchanged.

Again, this literature review was not exhaustive and does not contain all literature evaluated but excluded from model use. Some parts of the model were built around specific parameters that would not be retrievable in usable form without substantial redesign, like Mellors et al.'s and Gabillard et al.'s CD4 state transition probabilities (Mellors et al. 1997; Gabillard et al. 2013). But updates were made to treatment transition probabilities, mortality probabilities, initial CD4 count distribution among ART experienced, screening accuracy parameters, and ART costs for estimation.

REFERENCES

Ajose, Olawale, Siddharth Mookerjee, Edward J. Mills, Andrew Boulle, and Nathan Ford. 2012. "Treatment Outcomes of Patients on Second-Line Antiretroviral Therapy in Resource-Limited Settings: A Systematic Review and Meta-Analysis." *AIDS* 26 (8): 929–38. doi:10.1097/QAD.0b013e328351f5b2.

Arredondo, Miguel, Carolina Garrido, Neil Parkin, Natalia Zahonero, Silvia Bertagnolio, Vincent Soriano, and Carmen de Mendoza. 2012. "Comparison of HIV-1 RNA Measurements Obtained by Using Plasma and Dried Blood Spots in the Automated Abbott Real-Time Viral Load Assay." *Journal of Clinical Microbiology* 50 (3): 569–72. doi:10.1128/JCM.00418-11.

Balinda, Sheila N., Pascale Ondoa, Ekwaro A. Obuku, Aletta Kliphuis, Isaac Egau, Michelle Bronze, Lordwin Kasambula, et al. 2016. "Clinical Evaluation of an Affordable Qualitative Viral Failure Assay for HIV Using Dried Blood Spots in Uganda." *PloS One* 11 (1): e0145110. doi:10.1371/journal.pone.0145110.

Bangsberg, David R. 2008. "Preventing HIV Antiretroviral Resistance through Better Monitoring of Treatment Adherence." *The Journal of Infectious Diseases* 197 Suppl 3 (May): S272–78. doi:10.1086/533415.

Barth, Roos E., Maarten F. Schim van der Loeff, Rob Schuurman, Andy I. M. Hoepelman, and Annemarie M. J. Wensing. 2010. "Virological Follow-up of Adult Patients in Antiretroviral Treatment Programmes in Sub-Saharan Africa: A Systematic Review." *The Lancet Infectious Diseases* 10 (3): 155–66. doi:10.1016/S1473-3099(09)70328-7.

Bertagnolio, Silvia, Neil T. Parkin, Michael Jordan, James Brooks, and J. Gerardo García-Lerma. 2010. "Dried Blood Spots for HIV-1 Drug Resistance and Viral Load Testing: A Review of Current Knowledge and WHO Efforts for Global HIV Drug Resistance Surveillance." *AIDS Reviews* 12 (4): 195–208. http://www.ncbi.nlm.nih.gov/pubmed/21179184.

Boyer, Sylvie, Laura March, Charles Kouanfack, Gabrièle Laborde-Balen, Patricia Marino, Avelin Fobang Aghokeng, Eitel Mpoudi-Ngole, et al. 2013. "Monitoring of HIV Viral Load, CD4 Cell Count, and Clinical Assessment versus Clinical Monitoring Alone for Antiretroviral Therapy in Low-Resource Settings (Stratall ANRS 12110/ESTHER): A Cost-Effectiveness Analysis." *The Lancet Infectious Diseases* 13 (7): 577–86. doi:10.1016/S1473-3099(13)70073-2.

Calmy, Alexandra, Nathan Ford, Bernard Hirschel, Steven J. Reynolds, Lut Lynen, Eric Goemaere, Felipe Garcia de la Vega, Luc Perrin, and William Rodriguez. 2007. "HIV Viral Load Monitoring in Resource-Limited Regions: Optional or Necessary?" *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 44 (1): 128–34. doi:10.1086/510073.

Clinton Health Access Initiative. 2015. "ARV Market Report: The State of the Antiretroviral Drug Market in Low- and Middle-Income Countries, 2014-2019." Clinton Health Access Initiative. http://www.clintonhealthaccess.org/content/uploads/2015/11/CHAI-ARV-Market-Report-2015_FINAL.pdf.

Chaiwarith, Romanee, Charussri Wachirakaphan, Wilai Kotarathititum, Jutharat Praparatanaphan, Thira Sirisanthana, and Khuanchai Supparatpinyo. 2007.

"Sensitivity and Specificity of Using CD4+ Measurement and Clinical Evaluation to Determine Antiretroviral Treatment Failure in Thailand." *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases* 11 (5): 413–16. doi:10.1016/j.ijid.2006.11.003.

Cole, Stephen R., Sonia Napravnik, Michael J. Mugavero, Bryan Lau, Joseph J. Eron Jr, and Michael S. Saag. 2010. "Copy-Years Viremia as a Measure of Cumulative Human Immunodeficiency Virus Viral Burden." *American Journal of Epidemiology* 171 (2): 198–205. doi:10.1093/aje/kwp347.

Cori, Anne, Michael Pickles, Ard van Sighem, Luuk Gras, Daniela Bezemer, Peter Reiss, and Christophe Fraser. 2015. "CD4+ Cell Dynamics in Untreated HIV-1 Infection: Overall Rates, and Effects of Age, Viral Load, Sex and Calendar Time." *AIDS* 29 (18): 2435–46. doi:10.1097/QAD.0000000000000854.

Drabick, Joseph J., Wheaton J. Williams, Douglas B. Tang, Wellington Sun, Raymond C. Chung, and Military Medical Consortium For Applied Retroviral Research. 1992. "CD4 Lymphocyte Decline and Survival in Human Immunodeficiency Virus Infection." *AIDS Research and Human Retroviruses* 8 (12): 2039–47. doi:10.1089/aid.1992.8.2039.

Fajardo, Emmanuel, Carol A. Metcalf, Pascale Chaillet, Lucia Aleixo, Pieter Pannus, Isabella Panunzi, Laura Triviño, Tom Ellman, Andrew Likaka, and Reuben Mwenda. 2014. "Prospective Evaluation of Diagnostic Accuracy of Dried Blood Spots from Finger Prick Samples for Determination of HIV-1 Load with the NucliSENS Easy-Q HIV-1 Version 2.0 Assay in Malawi." *Journal of Clinical Microbiology* 52 (5): 1343–51. doi:10.1128/JCM.03519-13.

Ferreyra, Cecilia, Oliver Yun, Nell Eisenberg, Elena Alonso, Ashimosi S. Khamadi, Matilu Mwau, Martha Kihara Mugendi, et al. 2012. "Evaluation of Clinical and Immunological Markers for Predicting Virological Failure in a HIV/AIDS Treatment Cohort in Busia, Kenya." *PloS One* 7 (11): e49834. doi:10.1371/journal.pone.0049834.

Fox, Matthew P., Gilles Van Cutsem, Janet Giddy, Mhairi Maskew, Olivia Keiser, Hans Prozesky, Robin Wood, et al. 2012. "Rates and Predictors of Failure of First-Line Antiretroviral Therapy and Switch to Second-Line ART in South Africa." *Journal of Acquired Immune Deficiency Syndromes* 60 (4): 428–37. doi:10.1097/QAI.0b013e3182557785.

Gabillard, Delphine, Charlotte Lewden, Ibra Ndoye, Raoul Moh, Olivier Segeral, Besigin Tonwe-Gold, Jean-François Etard, et al. 2013. "Mortality, AIDS-Morbidity, and Loss to Follow-up by Current CD4 Cell Count Among HIV-1–Infected Adults Receiving Antiretroviral Therapy in Africa and Asia: Data From the ANRS 12222 Collaboration." *JAIDS Journal of Acquired Immune Deficiency Syndromes* 62 (5): 555–61. doi:10.1097/QAI.0b013e3182821821.

Gale, Howard B., Steven R. Gitterman, Heather J. Hoffman, Fred M. Gordin, Debra A. Benator, Ann M. Labriola, and Virginia L. Kan. 2013. "Is Frequent CD4+ T-Lymphocyte Count Monitoring Necessary for Persons with Counts >=300 cells/μL and HIV-1 Suppression?" *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 56 (9): 1340–43. doi:10.1093/cid/cit004.

Garrido, Carolina, Natalia Zahonero, Angélica Corral, Miguel Arredondo, Vincent Soriano, and Carmen de Mendoza. 2009. "Correlation between Human Immunodeficiency Virus Type 1 (HIV-1) RNA Measurements Obtained with Dried Blood Spots and Those Obtained with Plasma by Use of Nuclisens EasyQ HIV-1 and Abbott RealTime HIV Load Tests." *Journal of Clinical Microbiology* 47 (4): 1031–36. doi:10.1128/JCM.02099-08.

Girard, Pierre-Marie, Mark Nelson, Perry Mohammed, Andrew Hill, Yvon van Delft, and Christiane Moecklinghoff. 2013. "Can We Stop CD4+ Testing in Patients with HIV-1 RNA Suppression on Antiretroviral Treatment?" *AIDS* 27 (17): 2759–63. doi:10.1097/01.aids.0000432458.98851.c1.

Gupta, Ravindra K., Andrew Hill, Anthony W. Sawyer, Alessandro Cozzi-Lepri, Viktor von Wyl, Sabine Yerly, Viviane Dias Lima, Huldrych F. Günthard, Charles Gilks, and Deenan Pillay. 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): 409–17. doi:10.1016/S1473-3099(09)70136-7.

Hall, H. Irene, David R. Holtgrave, Tian Tang, and Philip Rhodes. 2013. "HIV Transmission in the United States: Considerations of Viral Load, Risk Behavior, and Health Disparities." *AIDS and Behavior* 17 (5): 1632–36. doi:10.1007/s10461-013-0426-z.

Hamers, Raph L., A. W. Sawyer, Martin Tuohy, Wendy S. Stevens, Tobias F. Rinke de Wit, Andrew M. Hill, and ART-A Consortium. 2012. "Cost-Effectiveness of Laboratory Monitoring for Management of HIV Treatment in Sub-Saharan

Africa: A Model-Based Analysis." *AIDS* 26 (13): 1663–72. doi:10.1097/QAD.0b013e3283560678.

Hyle, Emily P., Ilesh V. Jani, Jonathan Lehe, Amanda E. Su, Robin Wood, Jorge Quevedo, Elena Losina, 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. doi:10.1371/journal.pmed.1001725.

Hyle, Emily P., Paul E. Sax, and Rochelle P. Walensky. 2013. "Potential Savings by Reduced CD4 Monitoring in Stable Patients with HIV Receiving Antiretroviral Therapy." *JAMA Internal Medicine* 173 (18): 1746–48. doi:10.1001/jamainternmed.2013.9329.

Ingole, Nayana, Preeti Mehta, Amar Pazare, Supriya Paranjpe, and Purva Sarkate. 2013. "Performance of Immunological Response in Predicting Virological Failure." *AIDS Research and Human Retroviruses* 29 (3): 541–46. doi:10.1089/AID.2012.0266.

Institute for Health Metrics and Evaluation. 2015. "Access, Bottlenecks, Costs, and Equity: ABCE Project Cross-Country Protocol." Institute for Health Metrics and Evaluation.

Jourdain, Gonzague, Sophie Le Cœur, Nicole Ngo-Giang-Huong, Patrinee Traisathit, Tim R. Cressey, Federica Fregonese, Baptiste Leurent, et al. 2013. "Switching HIV Treatment in Adults Based on CD4 Count versus Viral Load Monitoring: A Randomized, Non-Inferiority Trial in Thailand." *PLoS Medicine* 10 (8): e1001494. doi:10.1371/journal.pmed.1001494.

Laprise, Claudie, Alexandra de Pokomandy, Jean-Guy Baril, Serge Dufresne, and Helen Trottier. 2013. "Virologic Failure Following Persistent Low-Level Viremia in a Cohort of HIV-Positive Patients: Results from 12 Years of Observation." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 57 (10): 1489–96. doi:10.1093/cid/cit529.

Laurent, Christian, Charles Kouanfack, Gabrièle Laborde-Balen, Avelin Fobang Aghokeng, Jules Brice Tchatchueng Mbougua, Sylvie Boyer, Maria Patrizia Carrieri, et al. 2011. "Monitoring of HIV Viral Loads, CD4 Cell Counts, and Clinical Assessments versus Clinical Monitoring Alone for Antiretroviral Therapy in Rural District Hospitals in Cameroon (Stratall ANRS 12110/ESTHER): A Randomised Non-Inferiority Trial." *The Lancet Infectious Diseases* 11 (11): 825–33. doi:10.1016/S1473-3099(11)70168-2.

Lofgren, Sarah M., Anne B. Morrissey, Caroline C. Chevallier, Anangisye I. Malabeja, Sally Edmonds, Ben Amos, David J. Sifuna, et al. 2009. "Evaluation of a Dried Blood Spot HIV-1 RNA Program for Early Infant Diagnosis and Viral Load Monitoring at Rural and Remote Healthcare Facilities." *AIDS* 23 (18): 2459–66. doi:10.1097/QAD.0b013e328331f702.

Madec, Yoann, Sandrine Leroy, Marie-Anne Rey-Cuille, Florence Huber, and Alexandra Calmy. 2013. "Persistent Difficulties in Switching to Second-Line ART in Sub-Saharan Africa — A Systematic Review and Meta-Analysis." Edited by Susan Marie Graham. doi:10.1371/journal.pone.0082724.

Maduna, Patrick H., Matt Dolan, Lwando Kondlo, Honey Mabuza, Judith N. Dlamini, Mike Polis, Thabo Mnisi, et al. 2015. "Morbidity and Mortality according to

Latest CD4+ Cell Count among HIV Positive Individuals in South Africa Who Enrolled in Project Phidisa." *PloS One* 10 (4): e0121843. doi:10.1371/journal.pone.0121843.

Medecins Sans Frontieres. 2014. "Getting to Undetectable: Usage of HIV Viral Load Monitoring in Five Countries." Medecins Sans Frontieres.

Medecins Sans Frontieres. 2016. "Untangling the Web of Antiretroviral Price Reductions." 18th ed. Medecins Sans Frontieres.

Mee, Paul, Katherine L. Fielding, Salome Charalambous, Gavin J. Churchyard, and Alison D. Grant. 2008. "Evaluation of the WHO Criteria for Antiretroviral Treatment Failure among Adults in South Africa." *AIDS* 22 (15): 1971–77. doi:10.1097/QAD.0b013e32830e4cd8.

Mellors, J. W., A. Muñoz, J. V. Giorgi, J. B. Margolick, C. J. Tassoni, P. Gupta, L. A. Kingsley, et al. 1997. "Plasma Viral Load and CD4+ Lymphocytes as Prognostic Markers of HIV-1 Infection." *Annals of Internal Medicine* 126 (12): 946–54. doi:10.7326/0003-4819-126-12-199706150-00003.

Mermin, Jonathan, John P. Ekwaru, Willy Were, Richard Degerman, Rebecca Bunnell, Frank Kaharuza, Robert Downing, et al. 2011. "Utility of Routine Viral Load, CD4 Cell Count, and Clinical Monitoring among Adults with HIV Receiving Antiretroviral Therapy in Uganda: Randomised Trial." *BMJ* 343 (November): d6792. doi:10.1136/bmj.d6792.

Monleau, Marjorie, Christelle Butel, Eric Delaporte, Francois Boillot, and Martine Peeters. 2010. "Effect of Storage Conditions of Dried Plasma and Blood Spots on HIV-1 RNA Quantification and PCR Amplification for Drug Resistance

Genotyping." *The Journal of Antimicrobial Chemotherapy* 65 (8): 1562–66. doi:10.1093/jac/dkq205.

Mugavero, Michael J., Sonia Napravnik, Stephen R. Cole, Joseph J. Eron, Bryan Lau, Heidi M. Crane, Mari M. Kitahata, et al. 2011. "Viremia Copy-Years Predicts Mortality among Treatment-Naive HIV-Infected Patients Initiating Antiretroviral Therapy." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 53 (9): 927–35. doi:10.1093/cid/cir526.

Napierala Mavedzenge, Sue, Calum Davey, Tarisai Chirenje, Phyllis Mushati, Sibongile Mtetwa, Jeffrey Dirawo, Boniface Mudenge, Andrew Phillips, and Frances M. Cowan. 2015. "Finger Prick Dried Blood Spots for HIV Viral Load Measurement in Field Conditions in Zimbabwe." *PloS One* 10 (5): e0126878. doi:10.1371/journal.pone.0126878.

Neogi, Ujjwal, Soham Gupta, Rashmi Rodridges, Pravat Nalini Sahoo, Shwetha D. Rao, Bharat B. Rewari, Suresh Shastri, Ayesha De Costa, and Anita Shet. 2011. "Dried Blood Spot HIV RNA Quantification: A Useful Tool for Viral Load Monitoring among HIVinfected Individuals in India." *Indian Journal of Medical Research* 136 (6): 956–62. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3612324/.

Phillips, Andrew N., Deenan Pillay, Alec H. Miners, Diane E. Bennett, Charles F. Gilks, and Jens D. Lundgren. 2008. "Outcomes from Monitoring of Patients on Antiretroviral Therapy in Resource-Limited Settings with Viral Load, CD4 Cell Count, or Clinical Observation Alone: A Computer Simulation Model." *The Lancet* 371 (9622): 1443–51. doi:10.1016/S0140-6736(08)60624-8.

Phillips, Andrew, Amir Shroufi, Lara Vojnov, Jennifer Cohn, Teri Roberts, Tom Ellman, Kimberly Bonner, et al. 2015. “Sustainable HIV Treatment in Africa through Viral-Load-Informed Differentiated Care.” *Nature* 528 (7580): S68–76. doi:10.1038/nature16046.

Rawizza, Holly E., Beth Chaplin, Seema T. Meloni, Geoffrey Eisen, Tara Rao, Jean-Louis Sankalé, Abdoulaye Dieng-Sarr, et al. 2011. “Immunologic Criteria Are Poor Predictors of Virologic Outcome: Implications for HIV Treatment Monitoring in Resource-Limited Settings.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 53 (12): 1283–90. doi:10.1093/cid/cir729.

Rutherford, George W., Andrew Anglemyer, Philippa J. Easterbrook, Tara Horvath, Marco Vitoria, Martina Penazzato, and Meg C. Doherty. 2014a. “Predicting Treatment Failure in Adults and Children on Antiretroviral Therapy: A Systematic Review of the Performance Characteristics of the 2010 WHO Immunologic and Clinical Criteria for Virologic Failure.” *AIDS* 28 Suppl 2 (March): S161–69. doi:10.1097/QAD.0000000000000236.

Rutstein, Sarah E., Mina C. Hosseinipour, Deborah Kamwendo, Alice Soko, Memory Mkandawire, Andrea K. Biddle, William C. Miller, et al. 2015. “Dried Blood Spots for Viral Load Monitoring in Malawi: Feasible and Effective.” *PloS One* 10 (4): e0124748. doi:10.1371/journal.pone.0124748.

Shen, Zhiyong, Qiuying Zhu, Zhenzhu Tang, Stephen W. Pan, Heng Zhang, He Jiang, Yi Chen, et al. 2016. “Effects of CD4 Cell Counts and Viral Load Testing on Mortality Rates in Patients With HIV Infection Receiving Antiretroviral

Treatment: An Observational Cohort Study in Rural Southwest China." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 63 (1): 108–14. doi:10.1093/cid/ciw146.

Siapka, Mariana, Michelle Remme, Carol Dayo Obure, Claudia B. Maier, Karl L. Dehne, and Anna Vassall. 2014. "Is There Scope for Cost Savings and Efficiency Gains in HIV Services? A Systematic Review of the Evidence from Low- and Middle-Income Countries." *Bulletin of the World Health Organization* 92 (7): 499–511AD. doi:10.2471/BLT.13.127639.

Smit, Pieter W., Kimberly A. Sollis, Susan Fiscus, Nathan Ford, Marco Vitoria, Shaffiq Essajee, David Barnett, et al. 2014. "Systematic Review of the Use of Dried Blood Spots for Monitoring HIV Viral Load and for Early Infant Diagnosis." *PloS One* 9 (3): e86461. doi:10.1371/journal.pone.0086461.

Templer, Stefanie P., Britta Seiverth, Paul Baum, Wendy Stevens, Carole Seguin-Devaux, and Sergio Carmona. 2016. "Improved Sensitivity of a Dual-Target HIV-1 Qualitative Test for Plasma and Dried Blood Spots." *Journal of Clinical Microbiology*, May. doi:10.1128/JCM.00128-16.

Tucker, Joseph D., Cedric H. Bien, Philippa J. Easterbrook, Meg C. Doherty, Martina Penazzato, Marco Vitoria, and Rosanna W. Peeling. 2014. "Optimal Strategies for Monitoring Response to Antiretroviral Therapy in HIV-Infected Adults, Adolescents, Children and Pregnant Women: A Systematic Review." *AIDS* 28 Suppl 2 (March): S151–60. doi:110.1097/QAD.0000000000000230.

UNAIDS. 2014. "90-90-90: An Ambitious Treatment Target to Help End the AIDS Epidemic." Joint United Nations Programme on HIV/AIDS.

UNAIDS. 2016. “Global AIDS Update 2016.” United Nations. http://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf.

Wood, Evan, Thomas Kerr, Greg Rowell, Julio S. G. Montaner, Peter Phillips, P. Todd Korthuis, and David L. Simel. 2014. “Does This Adult Patient Have Early HIV Infection?: The Rational Clinical Examination Systematic Review.” *JAMA: The Journal of the American Medical Association* 312 (3): 278–85. doi:10.1001/jama.2014.5954.

World Health Organization. 2014. “Technical and Operational Considerations for Implementing HIV Viral Load Testing.” World Health Organization.

World Health Organization. 2016. “Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection.” World Health Organization.

Table 1. Base case Markov model epidemiological parameters

Parameter	Value	Source
Treatment Transition Probabilities		
1st line ART failure, first 6 months	21%	Barth et al. 2010
1st line ART failure, after 6 months	2%	Fox et al. 2012
2nd line ART failure	8%	Ajose et al. 2012
CD4 State Transition Probabilities		
<100 to 100-199 cells/mm^3	45%	Gabillard et al. 2013
100-199 to 200-350 cells/mm^3	18%	Gabillard et al. 2013
200-350 to >350 cells/mm^3	10%	Gabillard et al. 2013
>350 to 200-350 cells/mm^3	22%	Mellors et al. 1997
200-350 to 100-199 cells/mm^3	22%	Mellors et al. 1997
100-199 to <100 cells/mm^3	31%	Mellors et al. 1997
CD4 Count Distribution Among ART Naïve		
CD4 >350 cells/mm^3	12%	IHME 2015*
CD4 200-350 cells/mm^3	30%	IHME 2015*
CD4 100-199 cells/mm^3	27%	IHME 2015*
CD4 <100 cells/mm^3	31%	IHME 2015*
CD4 Count Distribution Among ART Experienced		
CD4 >350 cells/mm^3	48%	Calibrated from ART Naïve**
CD4 200-350 cells/mm^3	36%	Calibrated from ART Naïve**
CD4 100-199 cells/mm^3	11%	Calibrated from ART Naïve**
CD4 <100 cells/mm^3	4%	Calibrated from ART Naïve**
Probability of Death		
CD4 >350 cells/mm^3	<1%	Maduna et al. 2015
CD4 200-350 cells/mm^3	1%	Maduna et al. 2015
CD4 100-199 cells/mm^3	2%	Maduna et al. 2015
CD4 <100 cells/mm^3	2%	Maduna et al. 2015

* Institute for Health Metrics and Evaluation (IHME)
** Initial distribution of CD4 states of the cohort of ART experienced patients was calibrated at entry to the CD4 count distribution of ART naïve patients who were on first-line ART and virologically suppressed after 5 years.

Table 2. Base case screening accuracy parameters

Parameter	Sensitivity	Specificity	Source
Screening Accuracy			
CSM	11%	90%	Rutherford et al. 2014
CD4	27%	86%	Rutherford et al. 2014
PVL	100%	100%	Assumed
DBS*	90%	90%	Assumed
DBS (Abbott)	95%	92%	Vojnov et al. 2014
DBS (Biocentric)	95%	55%	Vojnov et al. 2014
DBS (bioMérieux)	84%	95%	Vojnov et al. 2014
DBS (kPCR)	91%	88%	Vojnov et al. 2014
DBS (Roche, SPEX protocol)	99%	44%	Vojnov et al. 2014
DBS (Roche, FVE protocol)	81%	97%	Carmona and Mahlumba 2014

Table 3. Base case cost parameters (measured in USD)

Parameter	Cost	Source
Annual First-line ART Regimen per patient	$113	CHAI 2015
Annual Second-line ART Regimen per patient	$321	CHAI 2015
DBS Screen	$22	MSF 2014
PVL Screen	$25	MSF 2014
CD4 Screen	$24	Hyle et al. 2014
Clinic Visit	$20	Siapka et al. 2014

Table 4a. Number of deaths among 10,000 ART naive patients after 5 years of follow-up, varying sensitivity and specificity of the DBS viral load assay

		Sensitivity				
		100%	95%	90%	85%	80%
Specificity	100%	1590	1611	1634	1661	1692
	95%	1596	1616	1639	1666	1697
	90%	1611	1630	1653	1680	1711
	85%	1635	1655	1677	1703	1734
	80%	1669	1688	1709	1735	1764

Table 4b. Number of deaths among 10,000 ART experienced patients after 5 years of follow-up, varying sensitivity and specificity of the DBS viral load assay

		Sensitivity				
		100%	95%	90%	85%	80%
Specificity	100%	678	683	688	694	701
	95%	682	686	691	697	704
	90%	692	696	701	707	714
	85%	709	713	718	724	731
	80%	732	736	741	746	753

Table 4c. Number of downclassifications among 10,000 ART naive patients after 5 years of follow-up, varying sensitivity and specificity of the DBS viral load assay

		Sensitivity				
		100%	95%	90%	85%	80%
Specificity	100%	0	23	50	82	121
	95%	0	23	50	81	120
	90%	0	22	48	78	116
	85%	0	21	45	74	110
	80%	0	19	41	68	101

Table 4d. Number of downclassifications among 10,000 ART experienced patients after 5 years of follow-up, varying sensitivity and specificity of the DBS viral load assay

		Sensitivity				
		100%	95%	90%	85%	80%
Specificity	100%	0	32	68	111	161
	95%	0	31	68	110	159
	90%	0	30	66	107	154
	85%	0	29	62	101	147
	80%	0	26	57	94	137

Table 4e. Number of upclassifications among 10,000 ART naive patients after 5 years of follow-up, varying sensitivity and specificity of the DBS viral load assay

		Sensitivity				
		100%	95%	90%	85%	80%
Specificity	100%	0	0	0	0	0
	95%	62	62	62	62	62
	90%	243	243	243	243	243
	85%	528	528	528	528	528
	80%	895	895	895	895	895

Table 4f. Number of upclassifications among 10,000 ART experienced patients after 5 years of follow-up, varying sensitivity and specificity of the DBS viral load assay

		Sensitivity				
		100%	95%	90%	85%	80%
Specificity	100%	0	0	0	0	0
	95%	74	74	74	74	74
	90%	292	292	292	292	292
	85%	637	637	637	637	637
	80%	1082	1082	1082	1082	1082

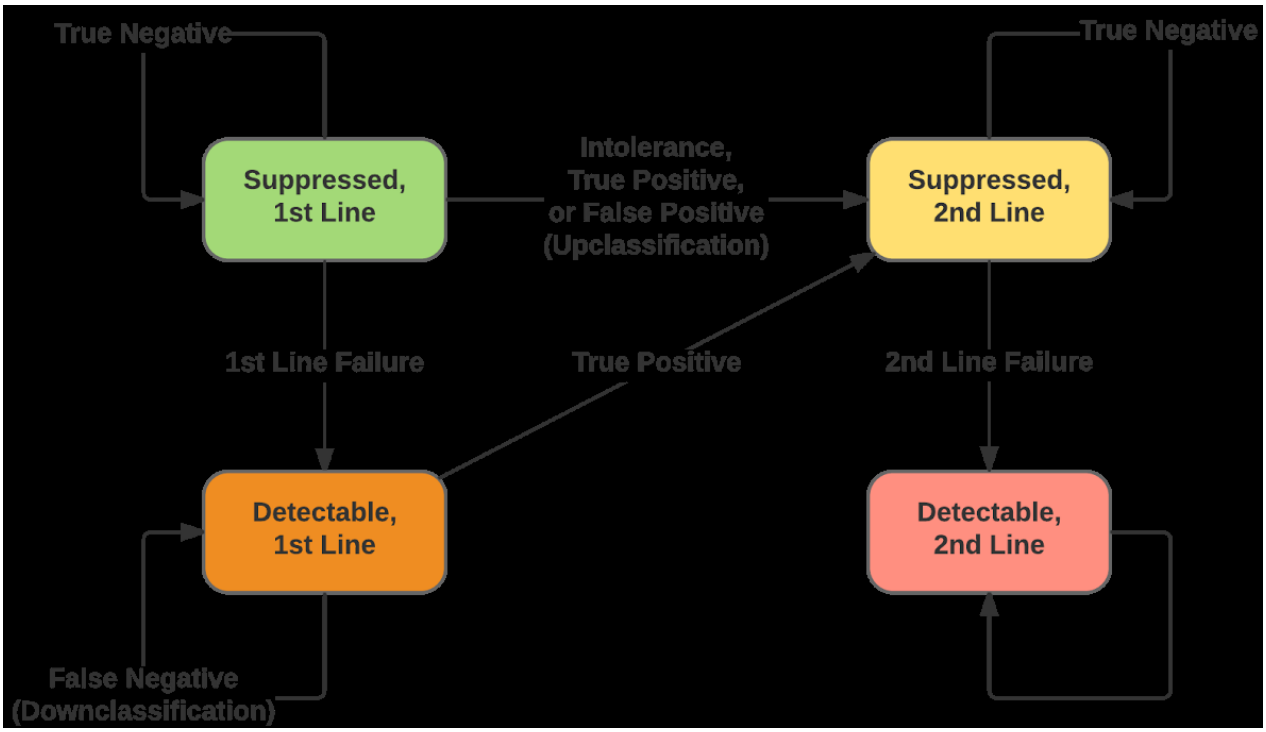
Table 4g. Overall costs for 10,000 ART naive patients after 5 years of follow-up, varying sensitivity and specificity of the DBS viral load assay (in millions of 2015 US dollars)

		Sensitivity				
		100%	95%	90%	85%	80%
Specificity	100%	$9.7	$9.7	$9.7	$9.6	$9.6
	95%	$9.7	$9.7	$9.7	$9.6	$9.6
	90%	$9.8	$9.8	$9.8	$9.8	$9.7
	85%	$10.0	$10.0	$10.0	$10.0	$9.9
	80%	$10.3	$10.3	$10.3	$10.2	$10.2

Table 4h. Overall costs for 10,000 ART experienced patients after 5 years of follow-up, varying sensitivity and specificity of the DBS viral load assay (in millions of 2015 US dollars)

		Sensitivity				
		100%	95%	90%	85%	80%
Specificity	100%	$9.0	$9.0	$9.1	$9.1	$9.1
	95%	$9.1	$9.1	$9.1	$9.1	$9.2
	90%	$9.2	$9.2	$9.3	$9.3	$9.3
	85%	$9.5	$9.5	$9.5	$9.5	$9.6
	80%	$9.8	$9.9	$9.9	$9.9	$9.9

Figure 1. Set of feasible Markov states and transitions as a result of virologic monitoring and failure



Note: this model assumes no third-line (or other) treatment is available, therefore no monitoring occurs for patients on second-line treatment.

Figure 2. Set of feasible Markov state combinations of virologic monitoring and disease progression

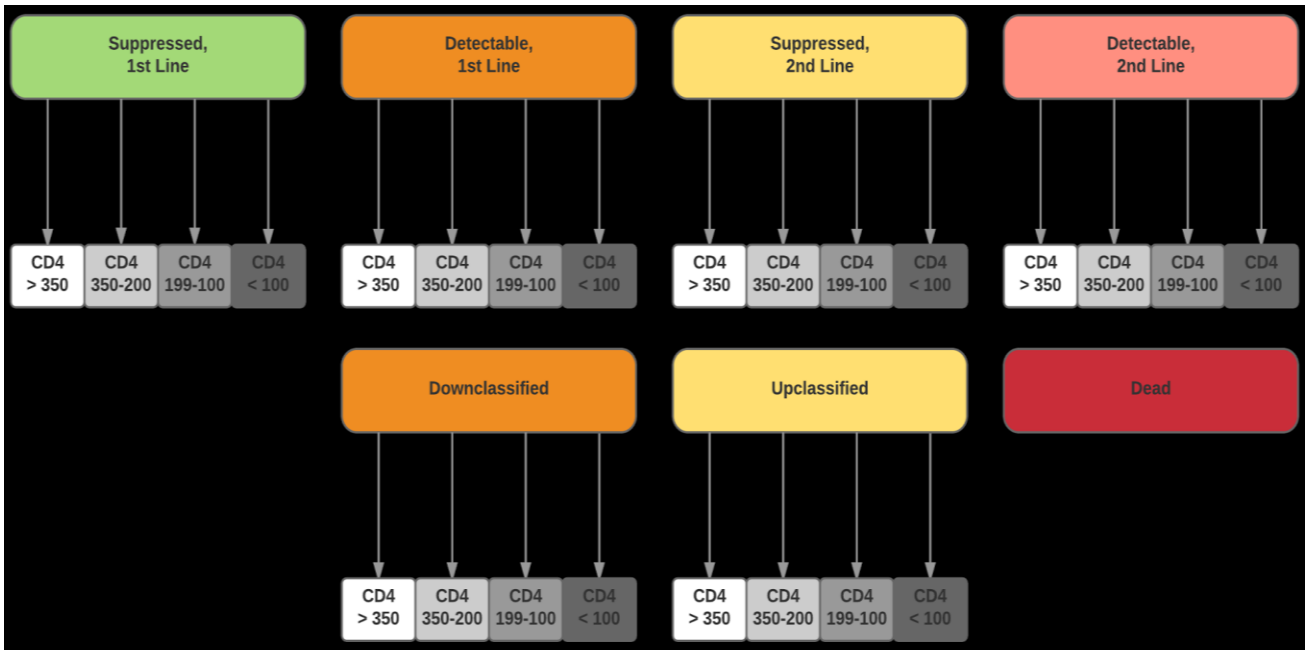
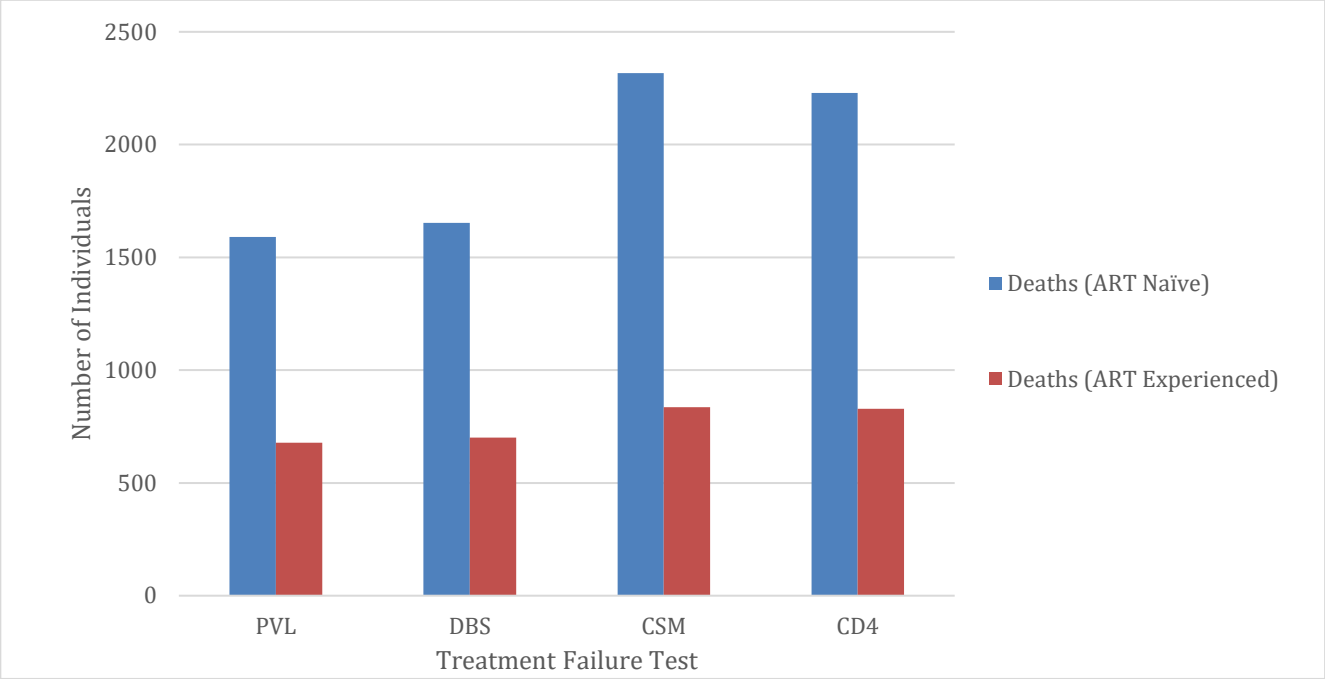
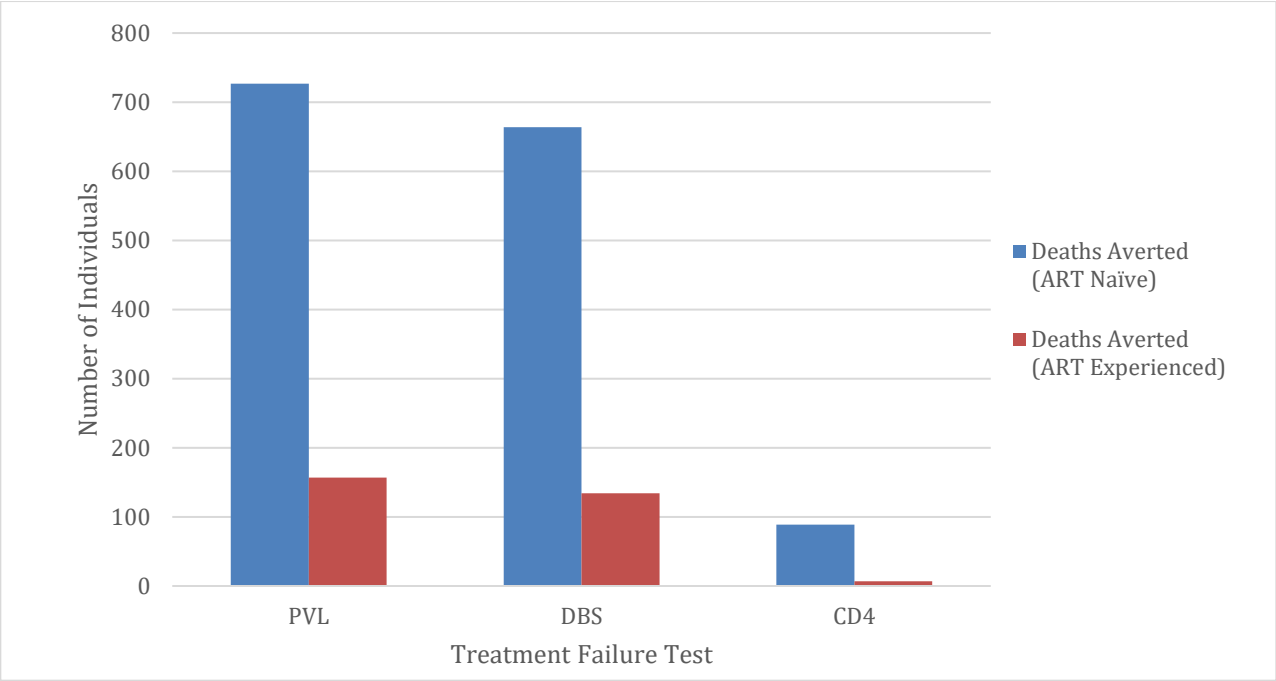


Figure 3a. Deaths among 10,000 ART naïve and experienced patients after 5 years of follow-up, by treatment failure test



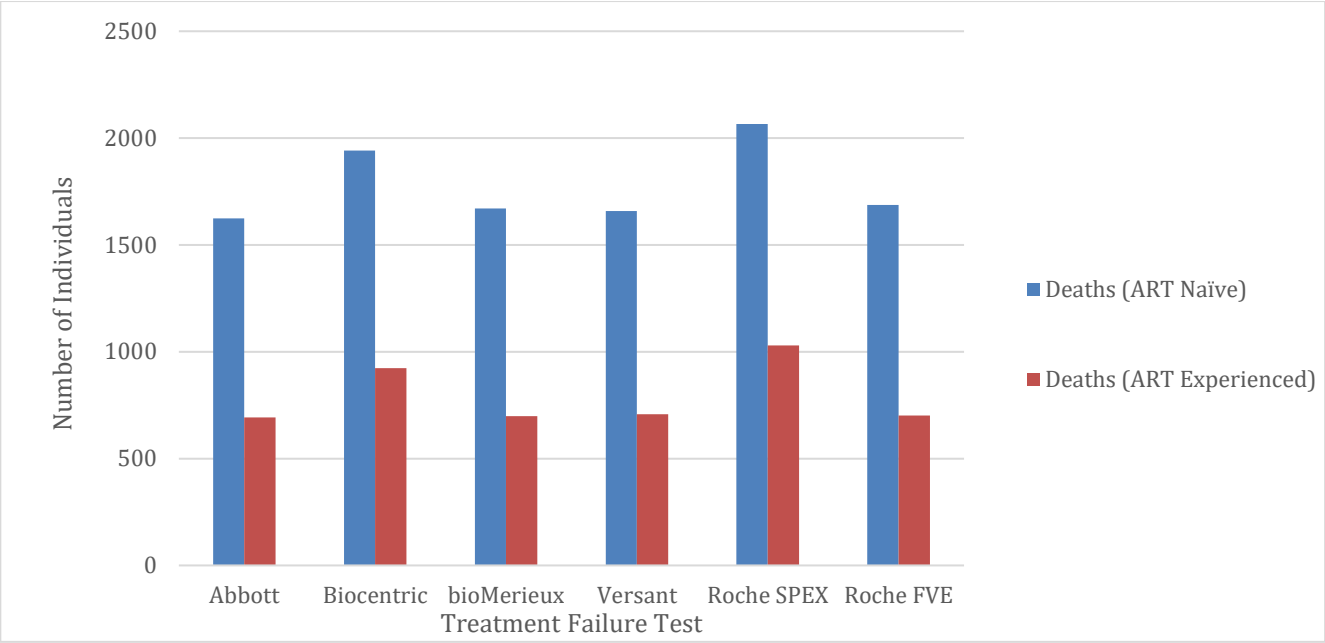
Screening Method	PVL	DBS	CSM	CD4
Sensitivity	100%	90%	11%	27%
Specificity	100%	90%	91%	86%
Deaths (ART Naïve)	1591	1653	2317	2229
Deaths (ART Experienced)	678	701	836	829

Figure 3b. Incremental deaths (compared with CSM) among 10,000 ART naïve and experienced patients after 5 years of follow-up, by treatment failure test



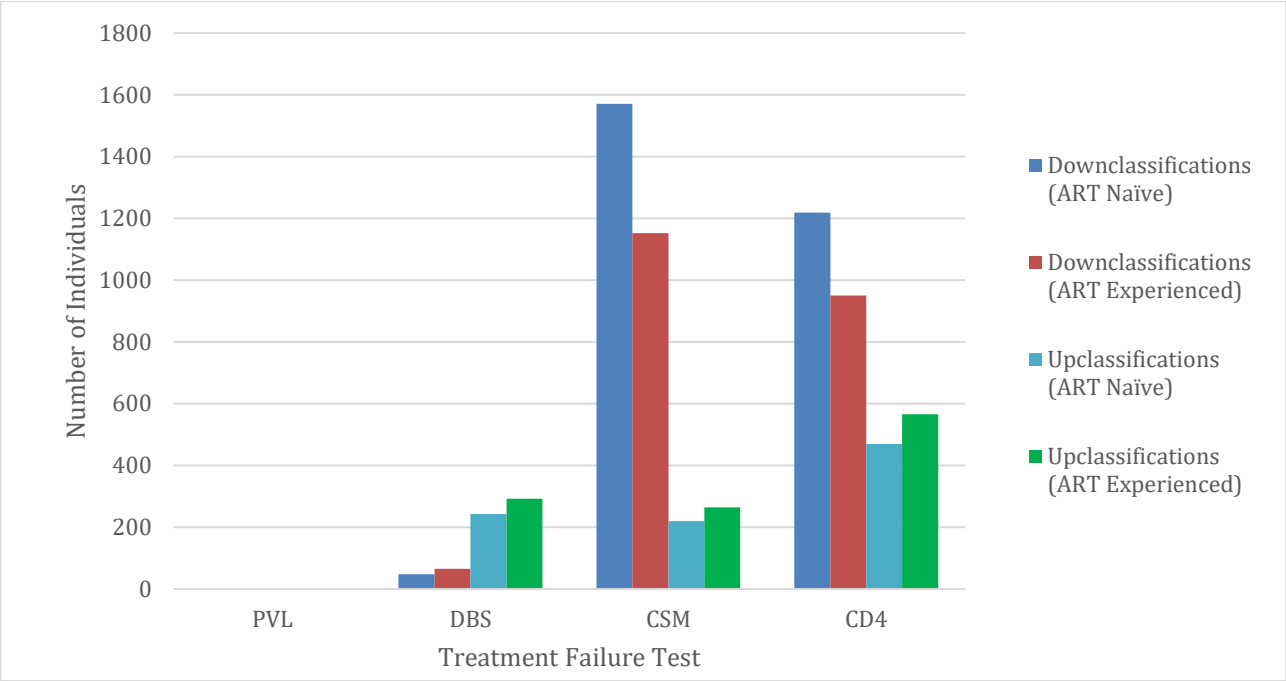
Screening Method	PVL	DBS	CD4
Sensitivity	100%	90%	27%
Specificity	100%	90%	86%
Deaths Averted (ART Naïve)	727	664	89
Deaths Averted (ART Experienced)	157	134	7

Figure 3c. Deaths among 10,000 ART naïve and experienced patients after 5 years of follow-up, by commercial DBS viral load assay



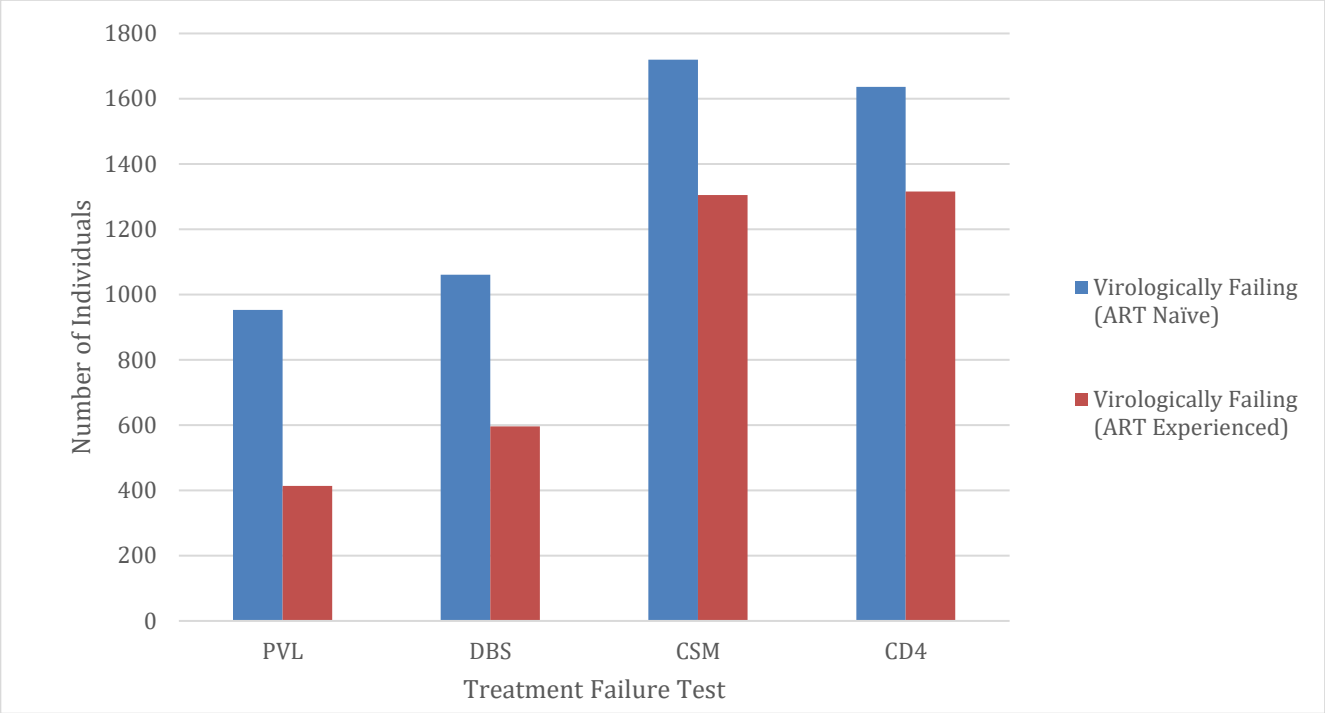
Screening Method	Abbott	Biocentric	bioMerieux	Versant	Roche SPEX	Roche FVE
Sensitivity	95%	95%	84%	91%	99%	81%
Specificity	92%	55%	95%	88%	44%	97%
Deaths (ART Naïve)	1623	1942	1670	1658	2065	1687
Deaths (ART Experienced)	692	923	699	707	1030	701

Figure 4a. Up- and downclassification among 10,000 ART naïve and experienced patients after 5 years of follow-up, by treatment failure test



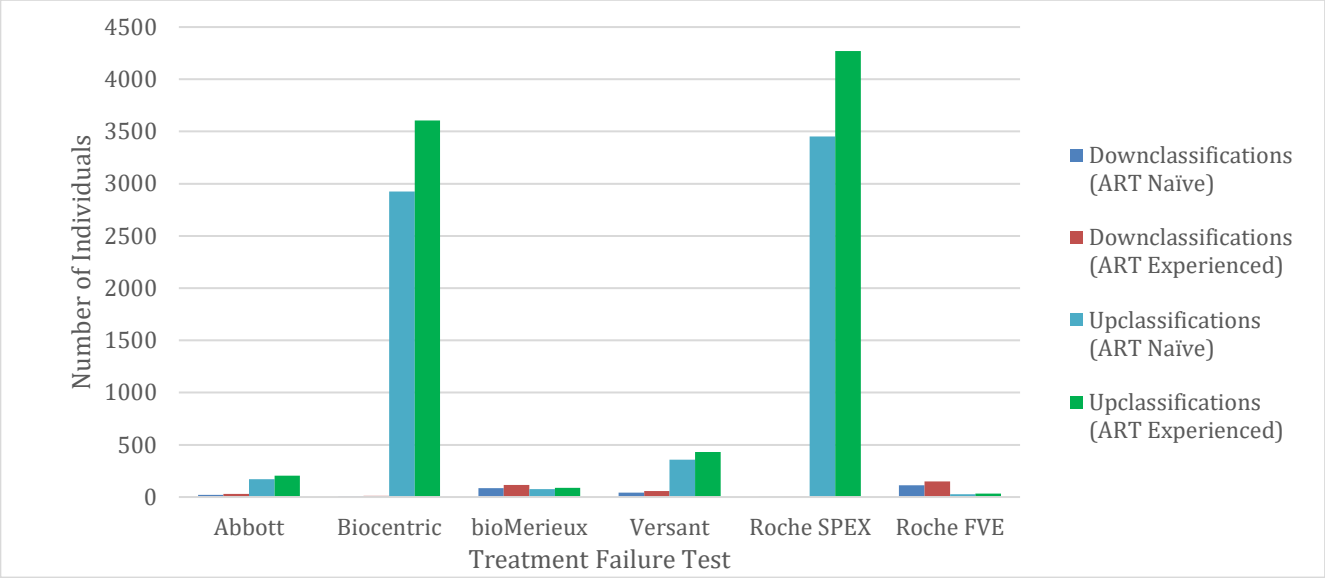
Screening Method	PVL	DBS	CSM	CD4
Sensitivity	100%	90%	11%	27%
Specificity	100%	90%	91%	86%
Downclassifications (ART Naïve)	0	48	1571	1218
Downclassifications (ART Experienced)	0	66	1152	950
Upclassifications (ART Naïve)	0	243	220	470
Upclassifications (ART Experienced)	0	292	264	566

Figure 4b. Number virologically failing (viral load above 1,000 copies/mL) among 10,000 ART naïve and experienced patients after 5 years of follow-up, by treatment failure test



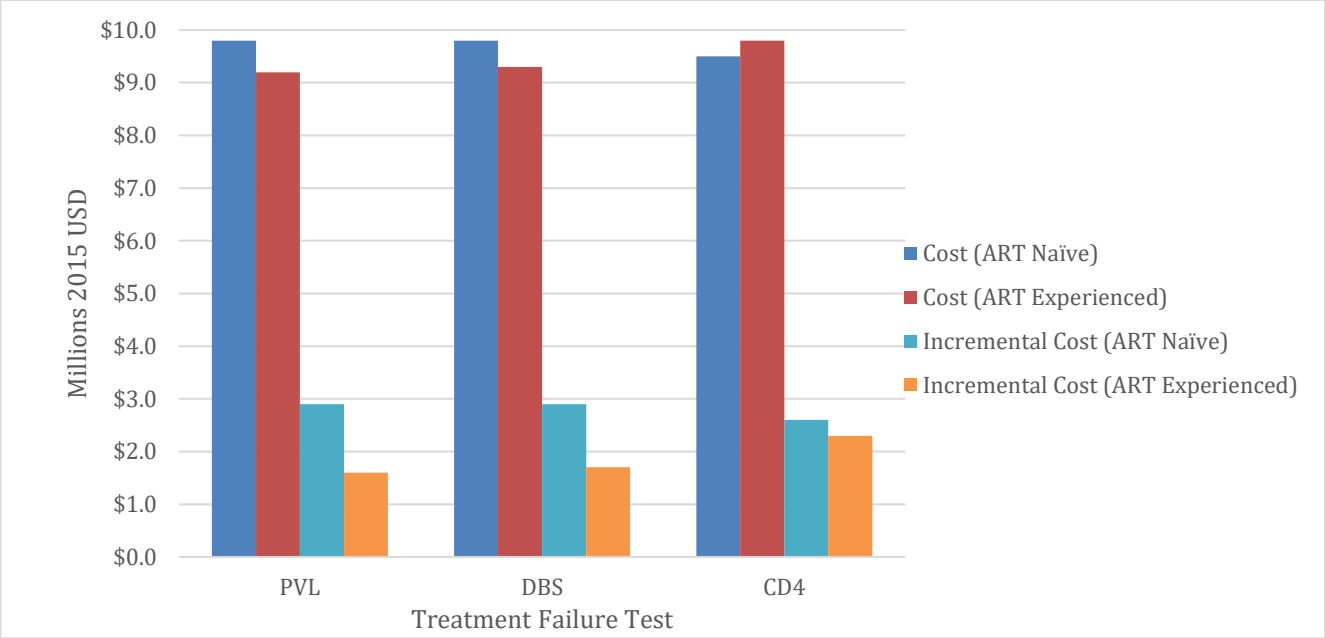
Screening Method	PVL	DBS	CSM	CD4
Sensitivity	100%	90%	11%	27%
Specificity	100%	90%	91%	86%
Virologically Failing (ART Naïve)	953	1061	1720	1636
Virologically Failing (ART Experienced)	414	596	1305	1316

Figure 4c. Up- and downclassification among 10,000 ART naïve and experienced patients after 5 years of follow-up, by commercial DBS viral load assay



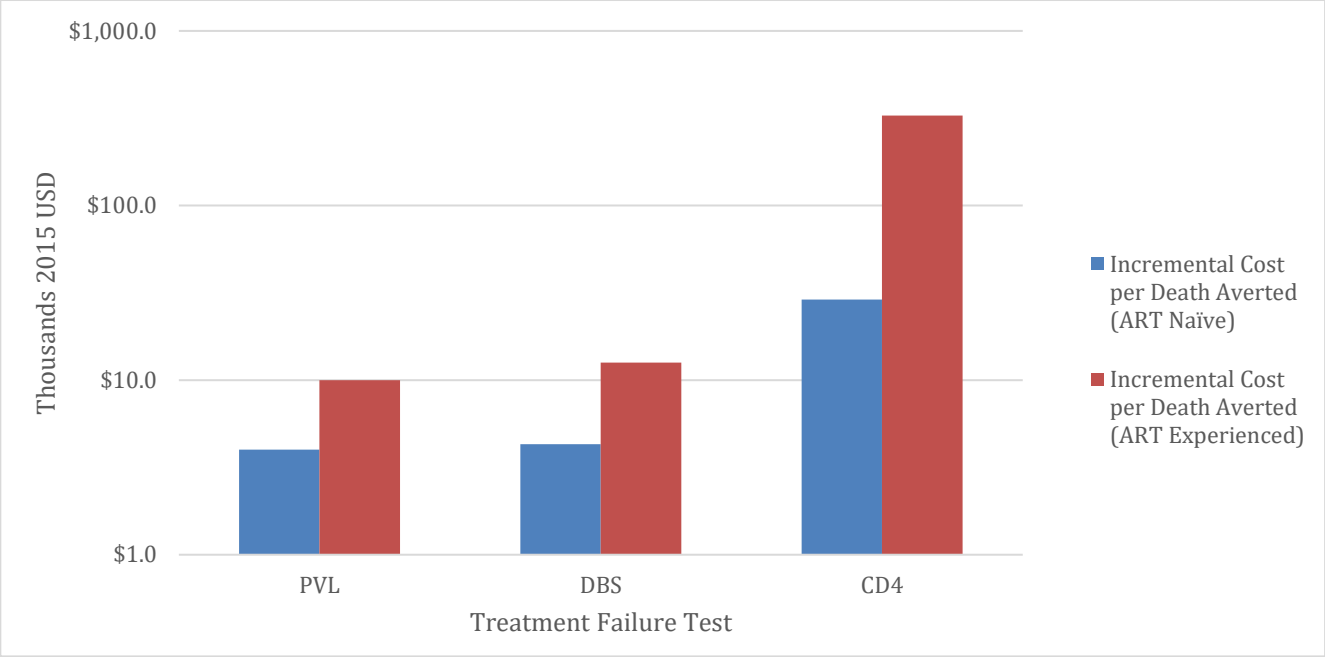
Screening Method	Abbott	Biocentric	bioMerieux	Versant	Roche SPEX	Roche FVE
Sensitivity	95%	95%	84%	91%	99%	81%
Specificity	92%	55%	95%	88%	44%	97%
Downclassifications (ART Naïve)	21	8	85	41	0	112
Downclassifications (ART Experienced)	29	12	116	57	1	149
Upclassifications (ART Naïve)	170	2926	74	359	3453	26
Upclassifications (ART Experienced)	204	3606	89	432	4270	32

Figure 5a. Cost and incremental cost (compared with CSM, in millions of 2015 US dollars) among 10,000 ART naïve and experienced patients over 5 years of follow-up, by treatment failure test



Screening Method	PVL	DBS	CD4
Sensitivity	100%	90%	27%
Specificity	100%	90%	86%
Cost (ART Naïve)	$9.8	$9.8	$9.5
Cost (ART Experienced)	$9.2	$9.3	$9.8
Incremental Cost (ART Naïve)	$2.9	$2.9	$2.6
Incremental Cost (ART Experienced)	$1.6	$1.7	$2.3

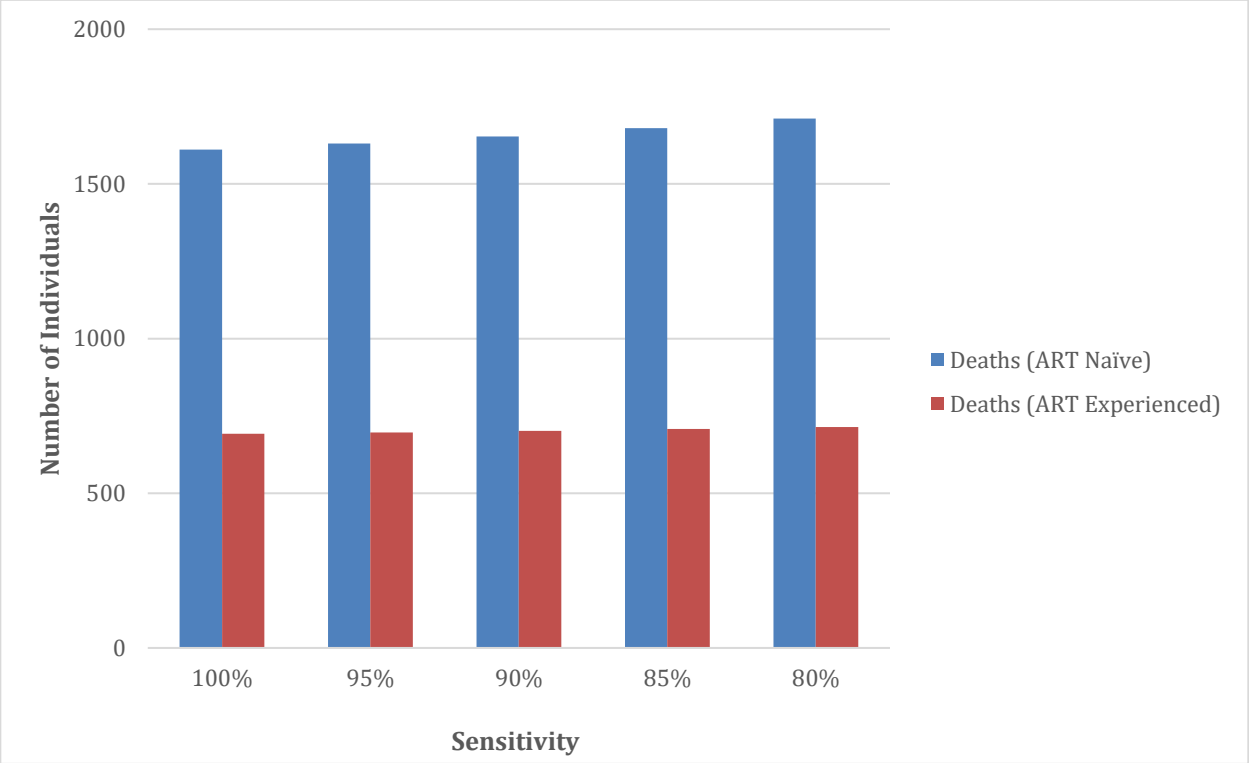
Figure 5b. Incremental cost per death averted (compared with CSM, in thousands of 2015 US dollars) among 10,000 ART naïve and experienced patients over 5 years of follow-up, by treatment failure test



Screening Method	PVL	DBS	CD4
Sensitivity	100%	90%	27%
Specificity	100%	90%	86%
Incremental Cost per Death Averted (ART Naïve)	$4.0	$4.3	$28.9
Incremental Cost per Death Averted (ART Experienced)	$10.0	$12.6	$328.0

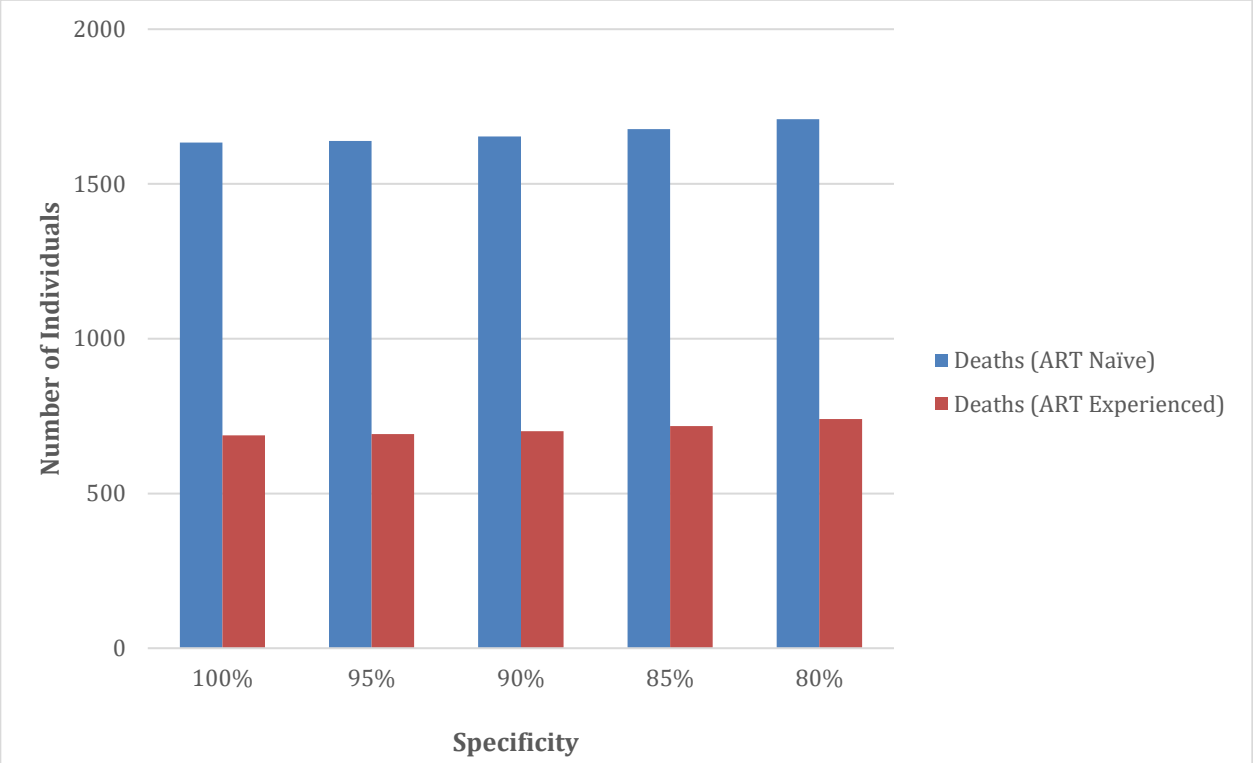
Note: incremental cost per death averted for was excessive in both cohorts for CD4, so the y-axis has been scaled logarithmically for in order to capture these values for readability.

Figure 6a. Deaths among 10,000 ART naïve and experienced patients after 5 years of follow-up, varying sensitivity of the DBS assay



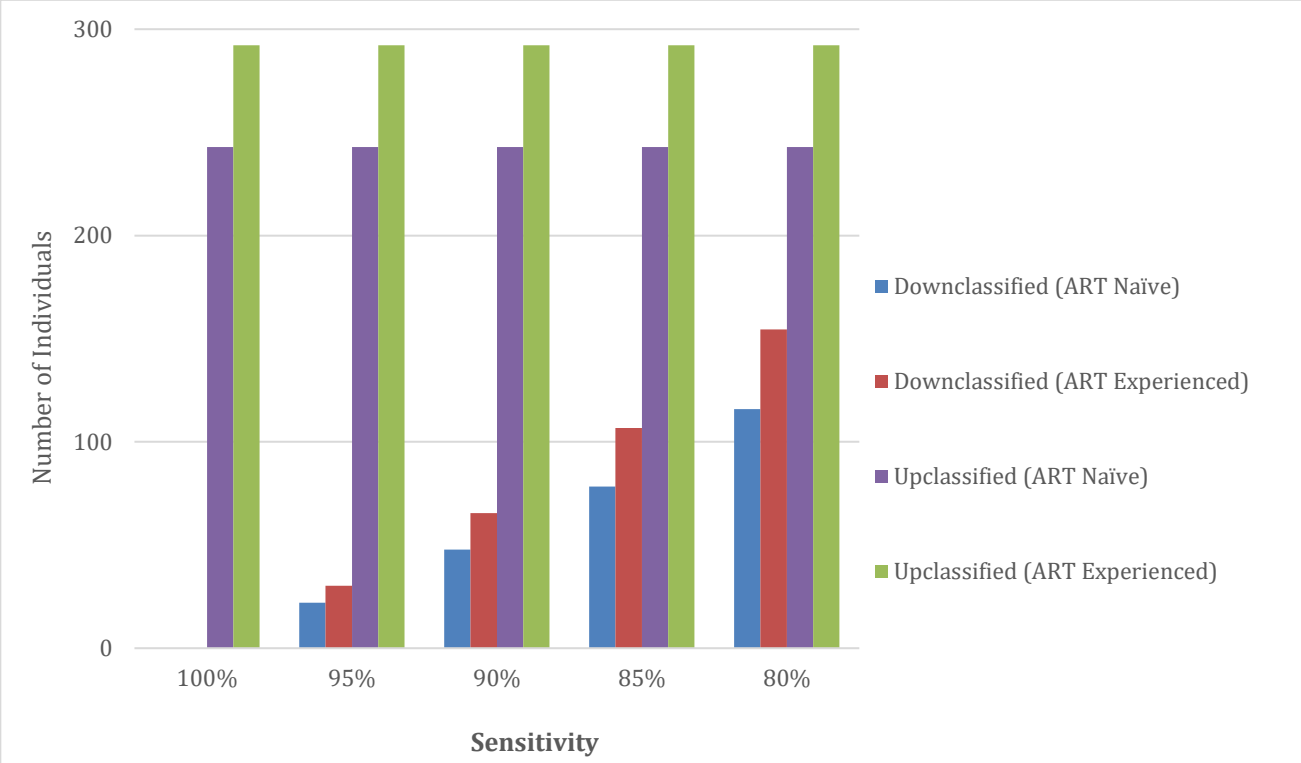
Sensitivity	100%	95%	90%	85%	80%
Specificity	90%	90%	90%	90%	90%
Deaths (ART Naïve)	1611	1630	1653	1680	1711
Deaths (ART Experienced)	692	696	701	707	714

Figure 6b. Deaths among 10,000 ART naïve and experienced patients after 5 years of follow-up, varying specificity of the DBS assay



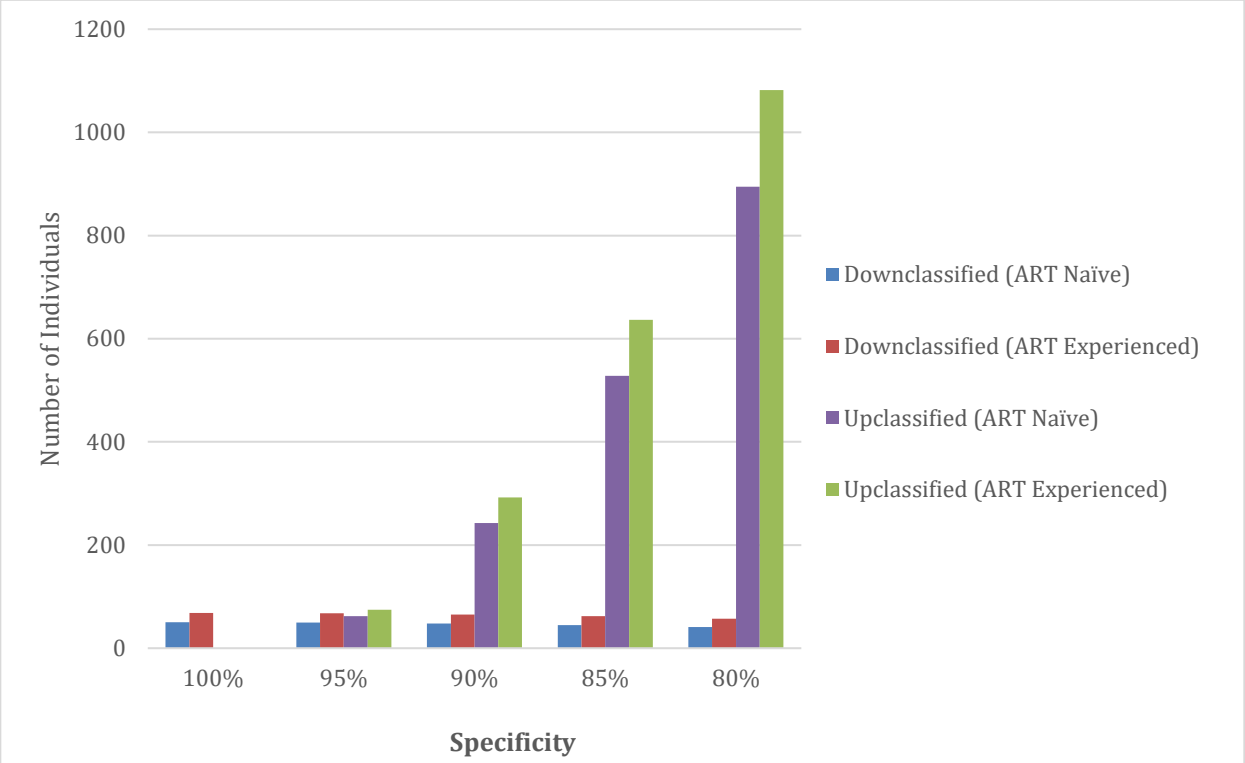
Sensitivity	90%	90%	90%	90%	90%
Specificity	100%	95%	90%	85%	80%
Deaths (ART Naïve)	1634	1639	1653	1677	1709
Deaths (ART Experienced)	688	691	701	718	741

Figure 6c. Up- and downclassification among 10,000 ART naïve and experienced patients after 5 years of follow-up, varying sensitivity of the DBS assay



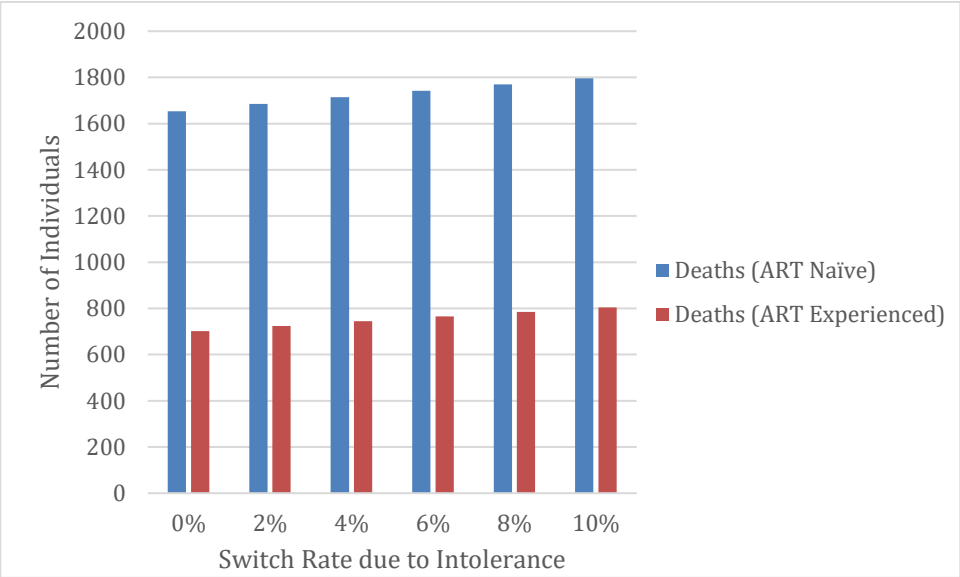
Sensitivity	100%	95%	90%	85%	80%
Specificity	90%	90%	90%	90%	90%
Downclassified (ART Naïve)	0	22	48	78	116
Downclassified (ART Experienced)	0	30	66	107	154
Upclassified (ART Naïve)	243	243	243	243	243
Upclassified (ART Experienced)	292	292	292	292	292

Figure 6d. Up- and downclassification among 10,000 ART naïve and experienced patients after 5 years of follow-up, varying specificity of the DBS assay



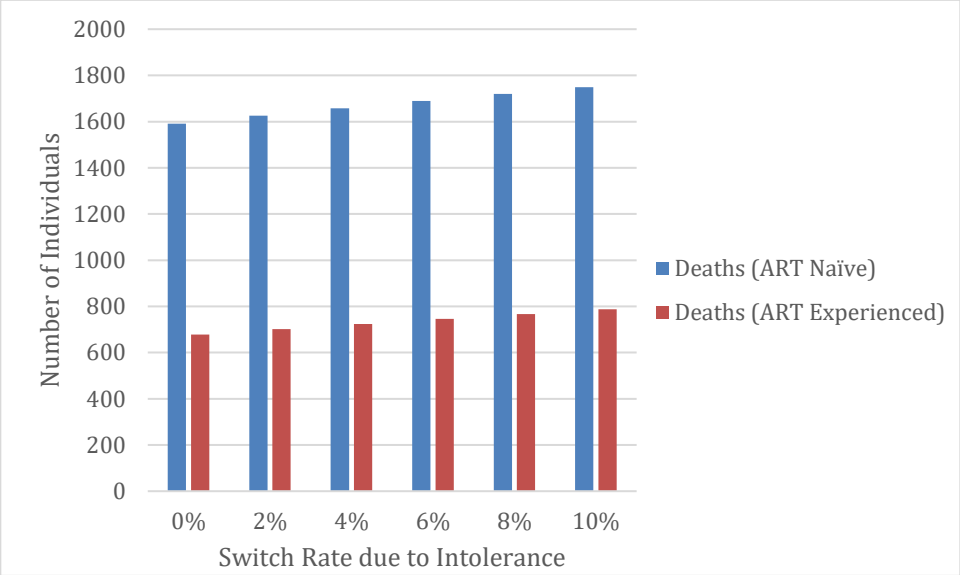
Sensitivity	90%	90%	90%	90%	90%
Specificity	100%	95%	90%	85%	80%
Downclassified (ART Naïve)	50	50	48	45	41
Downclassified (ART Experienced)	68	68	66	62	57
Upclassified (ART Naïve)	0	62	243	528	895
Upclassified (ART Experienced)	0	74	292	637	1082

Figure 7a. Deaths among 10,000 ART naïve and experienced patients after 5 years of follow-up, varying switch rate per follow-up period due to intolerance for DBS



Intolerance	0%	2%	4%	6%	8%	10%
Deaths (ART Naïve)	1653	1685	1714	1743	1770	1796
Deaths (ART Experienced)	701	723	745	765	785	805

Figure 7b. Deaths among 10,000 ART naïve and experienced patients after 5 years of follow-up, varying switch rate per follow-up period due to intolerance for PVL



Intolerance	0%	2%	4%	6%	8%	10%
Deaths (ART Naïve)	1591	1625	1658	1690	1720	1748
Deaths (ART Experienced)	678	702	724	746	767	787

Supplement Table T1. Base case Markov model epidemiological parameters

Parameter	Value	Source
Treatment Transition Probabilities		
1st line ART failure, first 6 months*	21%	Barth et al. 2010
1st line ART failure, after 6 months	2%	Fox et al. 2012
2nd line ART failure**	8%	Ajose et al. 2012
CD4 State Transition Probabilities		
<100 to 100-199 cells/mm^3	45%	Gabillard et al. 2013
100-199 to 200-350 cells/mm^3	18%	Gabillard et al. 2013
200-350 to >350 cells/mm^3	10%	Gabillard et al. 2013
>350 to 200-350 cells/mm^3	22%	Mellors et al. 1997
200-350 to 100-199 cells/mm^3	22%	Mellors et al. 1997
100-199 to <100 cells/mm^3	31%	Mellors et al. 1997
CD4 Count Distribution Among ART Naïve		
CD4 >350 cells/mm^3	12%	IHME 2015
CD4 200-350 cells/mm^3	30%	IHME 2015
CD4 100-199 cells/mm^3	27%	IHME 2015
CD4 <100 cells/mm^3	31%	IHME 2015
CD4 Count Distribution Among ART Experienced		
CD4 >350 cells/mm^3 †	48%	Calibrated from ART Naïve
CD4 200-350 cells/mm^3 †	36%	Calibrated from ART Naïve
CD4 100-199 cells/mm^3 †	11%	Calibrated from ART Naïve
CD4 <100 cells/mm^3 †	4%	Calibrated from ART Naïve
Probability of Death		
CD4 >350 cells/mm^3 ††	<1%	Maduna et al. 2015
CD4 200-350 cells/mm^3 ††	1%	Maduna et al. 2015
CD4 100-199 cells/mm^3 ††	2%	Maduna et al. 2015
CD4 <100 cells/mm^3 ††	2%	Maduna et al. 2015

***** Value changed from 18% to 21%. Source changed from Achieng et al. 2012 to Barth et al. 2010.
****** Value changed from 2% to 8%. It is unclear why this value was 2%.
† Values changed from 57%, 36%, 6%, and 1% to 48%, 36%, 11%, and 4% respectively, following updates to treatment transition probabilities.
†† Values changed from 1%, 1%, 2%, and 10% to <1%, 1%, 2%, 12% respectively. Source changed from Gabillard et al. 2013 to Maduna et al. 2015.

Supplement Table T2. Base case screening accuracy parameters

Parameter	Sensitivity	Specificity	Source
Screening Accuracy			
CSM	11%	90%	Rutherford et al. 2014
CD4	27%	86%	Rutherford et al. 2014
PVL	100%	100%	Assumed
DBS*	90%	90%	Assumed
DBS (Abbott)**	95%	92%	Vojnov et al. 2014
DBS (Biocentric)**	95%	55%	Vojnov et al. 2014
DBS (bioMérieux)**	84%	95%	Vojnov et al. 2014
DBS (kPCR)**	91%	88%	Vojnov et al. 2014
DBS (Roche, SPEX protocol)**	99%	44%	Vojnov et al. 2014
DBS (Roche, FVE protocol)**	81%	97%	Carmona and Mahlumba 2014

* Sensitivity changed from 95% to 90%, no change in specificity.
** Added for commercial test comparisons.

Supplement Table T3. Base case Markov model cost parameters

Parameter	Cost	Source
Annual First-line ART Regimen per patient *	$113	CHAI 2015
Annual Second-line ART Regimen per patient **	$321	CHAI 2015
DBS Screen	$22	MSF 2014
PVL Screen	$25	MSF 2014
CD4 Screen †	$24	Hyle et al. 2014
Clinic Visit ††	$20	Siapka et al. 2014

***** First-line regimen cost changed from $144 to $113. Source changed from MSF 2014 ("Unravelling the Web of ART Price Reductions") to CHAI 2015 ("ARV Market Report").
****** Second-line regimen cost changed from $288 to $321. Source changed from MSF 2014 ("Unravelling the Web of ART Price Reductions") to CHAI 2015 ("ARV Market Report").
† CD4 screen cost added.
†† Source changed to Siapka et al. 2014 from Phillips et al. 2015 to reflect primary (rather than secondary) source.

Supplement Literature Review S1. Treatment Failure Test Mortality Parameters

Authors	Journal, Year, (Study Period)	Type	Location, Size	Outcomes	Conclusions
Phillips et al.	Nature, 2015 (2015)	Simulated Cohort	Sub-Saharan Africa	Deaths per 100py: 5.85 with DBS, 5.53 with CD4, 7.5 with CSM	"Our results suggest that viral-load-informed differentiated ART care, using DBS sampling if necessary, is likely to be cost-effective in low-income settings in sub-Saharan Africa and is a sustainable model for providing ART."
Shen et al.	Clinical Infectious Diseases, 2015 (2008-2014)	Cohort	China, 39283	Adjusted mortality hazard ratio: 1.9 CSM vs CD4 (year 1), 1.0 CSM vs CD4 (year 3), 1.3 CSM vs PVL (year 1), 1.0 CSM vs PVL (year 3).	"Performing no VL tests in the first year after ART initiation was significantly associated with higher mortality rates. CD4 cell counts can be reduced to twice during the first year of ART and be reduced or stopped for patients who have achieved virologic suppression or immunologic stability after 12 months of treatment."
Maduna et al.	PLOS One, 2015 (2004-2008)	Multiple Cohort	South Africa, 7114	Mortality on ART by CD4 count per 100py: <100 (23.7), 100-199 (3.5), 200-349 (1.6), 350-499 (0.7), >500 (0.7)	"Rates of morbidity and mortality are lowest among those with CD4+ count of 350 or higher and rates do not differ for those with counts of 350–499 versus 500+ cells... Latest CD4+ specific death rates were greater for those not on ART as compared to those on ART."
Tucker et al.	AIDS, 2014 (up to 2013)	Systematic Review	Africa and Asia, 106439	Pooled mortality hazard ratio: 1.36 CSM vs CD4,	"CM + IM was shown to be beneficial in terms of a combined mortality and morbidity endpoint compared to CM alone. VM was associated with shorter duration of viremia and

				1.10-1.72 CD4 vs CD4 + PVL, 1.31-1.57 CSM vs CD4 + PVL	higher rates of switching, but an impact on mortality was not consistently shown."
Gabillard et al.	Journal of Acquired Immune Deficiency Syndromes, 2013 (1996-2009)	Multiple Cohort	Africa and Asia, 3917	Mortality by CD4 count per 100py: <50 (20.6), 51-100 (11.8), 101-200 (6.7), 201-350 (3.3), 351-500 (1.8), 501-650 (0.9), >650 (0.3) CD4 increase by CD4 count after ART initiation per mL: <50 (245), >200 (198)	"Death and AIDS rates remained substantial after ART initiation, even in individuals with high CD4 cell counts. Death rates appeared comparable with those observed in Western countries while AIDS rates appeared higher."
Boyer et al.	Lancet Infectious Diseases, 2013 (2006-2008)	Randomized Clinical Trial	Cameroon, 385	Deaths per 100py: 1.8 with PVL + CD4, 4.7 with CSM	"Laboratory monitoring is not cost effective in Cameroon compared with clinical monitoring alone in the current context of fixed health-care budgets. The cost-effectiveness of the laboratory strategy strongly improved in patients starting ART with a CD4 cell count of fewer than 200 cells per μL, because more life-years are saved."
Jourdain et al.	PLOS Medicine, 2013 (2005-2007)	Randomized Clinical Trial	Thailand, 2013	Deaths per 100py: 1.4 with PVL, 1.1 with CD4	"A CD4-based switching strategy was non-inferior in terms of clinical outcomes at 3 years of follow-up, compared to a reference VL based switching strategy. Moreover, at study end there were no differences in terms of viral suppression and immune restoration between arms."

Mermin et al.	British Medical Journal, 2011 (2003-2007)	Randomized Clinical Trial	Uganda, 1094	Adjusted mortality hazard ratio: 1.57 CSM vs PVL, 1.43 CSM vs CD4, 1.10 CD4 vs PVL	"People in the CD4 cell count monitoring arm did better than those with clinical monitoring alone... We did not find that quarterly viral loads provided additional clinical benefit to patients."
Laurent et al.	Lancet Infectious Diseases, 2011 (2006-2008)	Randomized Clinical Trial	Cameroon, 459	Mortality hazard ratio: 1.31 CSM vs CD4	"Clinical monitoring alone is not non-inferior to clinical monitoring plus laboratory monitoring in terms of mean increase in CD4 cell count to 2 years... Survival at 24 months was very similar between groups."
Phillips et al.	Lancet, 2008 (2008)	Simulated Cohort	Africa	Deaths per 100py: 9.7 with CSM, 8.0 with PVL	"For patients on the first-line regimen… the benefits of viral load or CD4 cell count monitoring over clinical monitoring alone are modest. Development of cheap and robust versions of these assays is important, but widening access to antiretrovirals—with or without laboratory monitoring—is currently the highest priority."

Supplement Literature Review S2. Treatment Failure Test Accuracy Parameters

Authors	Journal, Publication Year (Study Period)	Type	Location, Size	Outcomes	Conclusions
Mavedzenge et al.	PLOS One, 2015 (2014-2015)	Community Randomized Trial	Zimbabwe, 149	Accuracy: sensitivity 0.874, specificity 0.968 (bioMerieux)	"There was generally good agreement between DBS and plasma VL for detection of VL>1000. Overall, finger prick DBS appeared to be an acceptable sample for classifying VL as above or below 1000 copies/mL using the NucliSENS assay."
Vojnov et al.	World Health Organization, 2014 (2014)	Systematic Review	Africa	Pooled accuracy: sensitivity 0.95, specificity 0.92 (Abbott); sensitivity 0.95, specificity 0.55 (Biocentric); sensitivity 0.84, specificity 0.95 (bioMerieux); sensitivity 0.99, specificity 0.44 (Roche SPEX); sensitivity 0.91, specificity 0.88 (Versant)	"Some of the assay types were found to have low sensitivity at the time this evaluation was performed (up to June 2014) and should be avoided. While this reduced sensitivity means that plasma specimens are preferred for viral load testing, modelling suggests that if viral load testing with DBS specimens can be performed with reasonable sensitivity and specificity (>85%), then costs and outcomes are similar." *Note: both this and the following study by Carmona and Mahlumba have been cited in multiple WHO reports, but are unpublished. The true date of the final study report appears to be 2014, but has been cited by the WHO as 2016. The results cited in both years were identical.*
Carmona and Mahlumba	World Health Organization, 2014 (2014)	Systematic Review	Africa	Pooled accuracy: sensitivity 0.85, specificity 0.94 (Roche FVE)	"Some of the assay types were found to have low sensitivity at the time this evaluation was performed (up to June 2014) and should be avoided. While this reduced sensitivity means that plasma specimens are preferred

					for viral load testing, modelling suggests that if viral load testing with DBS specimens can be performed with reasonable sensitivity and specificity (>85%), then costs and outcomes are similar."
Rutherford et al.	AIDS, 2014 (2006-2012)	Systematic Review	Africa and South America, 15581	Pooled accuracy: sensitivity 0.110, specificity 0.905 with CSM; sensitivity 0.266, specificity 0.859 with CD4	"The 2010 WHO clinical and immunologic criteria are insensitive and have low PPV for predicting virologic failure. These data support the strong recommendation 2013 treatment guidelines that viral load testing be used to monitor for, diagnose, and confirm ART failure."
Ferreyra et al.	PLOS One, 2012 (2008)	Cross-sectional Cohort	Kenya, 926	Accuracy: sensitivity 0.364, specificity 0.835 with CSM + CD4	"Clinical and immunological criteria were found to perform relatively poorly in predicting virological failure of ART. VL monitoring and new algorithms for assessing clinical or immunological treatment failure, as well as improved adherence strategies, are required in ART programs in resource-limited settings."
Rawizza et al.	Clinical Infectious Diseases, 2011 (2004-2008)	Cohort	Nigeria, 9690	Accuracy: sensitivity 0.584, specificity 0.750 for CD4	"This analysis, which uses programmatic treatment data, a large sample size, longer duration of follow-up, and rates of virologic failure commensurate with other real-life treatment cohorts in RLS, provides the strongest evidence to date of the poor performance of immunologic criteria in identifying treatment failure."
Ingole et al.	AIDS Research and Human Retroviruses, 2013 (2008-2010)	Cohort	India, 130	Accuracy: sensitivity 0.111-0.400, specificity 0.298-0.925 for CD4	"Immunological criteria do not accurately predict virological failure resulting in significant misclassification of therapeutic responses. There is an urgent need for inclusion of viral load testing in the initiation and monitoring of ART."
Mee et al.	AIDS, 2008 (2002-2004)	Cohort	South Africa, 324	Accuracy: sensitivity 0.333,	"WHO clinical and CD4 criteria have poor sensitivity and specificity in detecting virological failure… Individuals with adequate virological suppression risk being

				specificity 0.856 for CD4; sensitivity 0.152, specificity 0.881 for CSM	incorrectly classified as having treatment failure and unnecessarily switched to second-line therapy. Virological failure should be confirmed before switching to second-line therapy."
Chaiwarith et al.	International Journal of Infectious Diseases, 2007 (2003-2005)	Retrospective Cohort	Thailand, 327	Accuracy: sensitivity 0.100, specificity 0.956 for CSM; sensitivity 0.200, specificity 0.859 for CD4.	"Our study, which was limited by small numbers, was not able to demonstrate that immunological or clinical criteria can adequately replace virological criteria for the determination of treatment failure."

BIOGRAPHICAL STATEMENT

Zach Adams was born in 1992 at The Johns Hopkins Hospital in Baltimore, Maryland. He obtained his B.A. from Franklin & Marshall College in Lancaster, Pennsylvania in 2014 with departmental honors in Public Health for his thesis on "Health Survey of the Lancaster County Plain Communities and Propensity Score Matching for Asthma in Amish Women" under the advisory of Kirk Miller. Upon graduation, he moved back to Baltimore to attend The Johns Hopkins Bloomberg School of Public Health. Zach's favorite hobby is photography, which he enjoys with his partner, Helen. He will be moving to Washington DC in the Fall to continue his career in health sciences.