USAID
FROM THE AMERICAN PEOPLE

ASSESSING THE IMPACT OF IMPERFECT IMMUNOASSAYS ON HIV PREVALENCE ESTIMATES FROM SURVEYS CONDUCTED BY THE DHS PROGRAM

DHS METHODOLOGICAL REPORTS 22

SEPTEMBER 2017

This publication was produced for review by the United States Agency for International Development. It was prepared by Mathieu Maheu-Giroux, Lawrence Joseph, Patrick Belisle, Samantha Lancione, and Jeffrey W. Eaton.

DHS Methodological Report No. 22

Assessing the Impact of Imperfect Immunoassays on HIV Prevalence Estimates from Surveys Conducted by The DHS Program

Mathieu Maheu-Giroux

Lawrence Joseph

Patrick Belisle

Samantha Lancione

Jeffrey W. Eaton

ICF

Rockville, Maryland, USA

September 2017

*Corresponding author: Mathieu Maheu-Giroux, McGill University, 1020 Avenue des Pins Ouest, Montréal, Québec, Canada; phone : (+1) 514-398-5110; email : mathieu.maheu-giroux@mcgill.ca

Acknowledgment: We are grateful for Joy Fishel, Shireen Assaf, and Dean Garrett for their helpful suggestions and comments on an earlier version of this report.

Editor: Diane Stoy
Document Production: Joan Wardell

This study was conducted with support provided by the United States Agency for International Development (USAID) through The DHS Program (#AID-OAA-C-13-00095). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

Recommended citation:

Maheu-Giroux, Mathieu, Lawrence Joseph, Patrick Belisle, Samantha Lancione, and Jeffrey W Eaton. 2017. *Assessing the Impact of Imperfect Immunoassays on HIV Prevalence Estimates from Surveys Conducted by The DHS Program.* DHS Methodological Reports No. 22. Rockville, Maryland, USA: ICF.

Contents

Tables

Figures

Preface

The Demographic and Health Surveys (DHS) Program is one of the principal sources of international data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services.

One of the objectives of The DHS Program is to continually assess and improve the methodology and procedures used to carry out national-level surveys as well as to offer additional tools for analysis. Improvements in methods used will enhance the accuracy and depth of information collected by The DHS Program and relied on by policymakers and program managers in low- and middle-income countries.

While data quality is a main topic of the DHS Methodological Reports series, the reports also examine issues of sampling, questionnaire comparability, survey procedures, and methodological approaches. The topics explored in this series are selected by The DHS Program in consultation with the U.S. Agency for International Development.

It is hoped that the DHS Methodological Reports will be useful to researchers, policymakers, and survey specialists, particularly those engaged in work in low- and middle-income countries, and will be used to enhance the quality and analysis of survey data.

Sunita Kishor
Director, The DHS Program

Abstract

The DHS Program has supported numerous large-scale HIV seroprevalence surveys, some of which used a testing strategy based on enzyme-immunoassays (EIA). There have been recent concerns that this algorithm could have led to overestimation of HIV prevalence. This report investigates the impact of potential misclassification of samples on HIV prevalence estimates for 23 surveys conducted from 2010-2014. In addition to visual inspection of laboratory results, we examined how accounting for potential misclassification of HIV status through Bayesian latent class models affected the prevalence estimates. Two types of Bayesian models were specified: a model that only uses the individual dichotomous test results and a continuous model that uses the quantitative information of the EIA (i.e., the signal-to-cutoff values). Overall, we found that adjusted prevalence estimates matched the surveys' original results, with overlapping uncertainty intervals. This suggested that misclassification of HIV status should not affect the prevalence estimates in most surveys. However, our analyses suggested that two surveys may be problematic. The prevalence could have been overestimated in the Uganda *AIDS Indicator Survey* 2011 and the Zambia *Demographic and Health Survey* 2013-14, although the magnitude of overestimation remains difficult to ascertain. Interpreting results from the Uganda survey is difficult because of the lack of internal quality control and potential violation of the multivariate normality assumption of the continuous Bayesian latent class model. In conclusion, despite the limitations of our latent class models, our analyses suggest that prevalence estimates from most of the surveys reviewed are not affected by sample misclassification.

KEY WORDS: HIV, enzyme immunoassays, surveillance, seroprevalence survey, Bayesian latent class models.

Executive Summary

There have been concerns about HIV laboratory testing results from surveys conducted by *The DHS Program* prior to 2015. The enzyme-immunoassay (EIA) based testing strategies used in these surveys could have led, in some instances, to an overestimation of HIV prevalence. These concerns led *The DHS Program* to conduct a review of the performance of these EIAs in 20 surveys conducted between 2010 and 2014.

This report extends the preceding analyses of laboratory data by assessing the impact of potential serological misclassification of the HIV prevalence estimates from 23 surveys (20 of these were previously reviewed by *The DHS Program*). Along with visual inspection of laboratory results, we accounted for the uncertainty associated with potential misclassification of EIA results and confirmatory assays by using two different Bayesian latent class models. The first model used dichotomous test results, while the second used the continuous information on the distribution of signal-to-cutoff values of the EIA.

Overall, the prevalence estimates from the latent class models roughly matched the survey results, with overlapping uncertainty intervals. This suggested that the impact of misclassification of HIV status is relatively minor in most surveys conducted by *The DHS Program* during 2010-2014. However, the results from the EIA-based testing algorithms used in the Uganda *AIDS Indicator Survey* 2011 and Zambia *Demographic and Health Survey* 2013-14 are potentially biased. The precise magnitude of this bias remains difficult to ascertain, especially for Uganda. In the Uganda survey, the lack of internal quality control and probable violation of the multivariate normality assumption of the continuous Bayesian latent class model limits our confidence in results from this adjustment. In this survey, the true unobserved HIV prevalence is probably closer to results from the rapid diagnostic test algorithm, which was conducted during the survey visit to the household, rather than the EIA-based algorithm, as suggested by our examination of HIV trends by age and sex. In the case of Zambia, the EIA-based algorithm produced a high number of weakly reactive samples, especially among men, which were later found to be negative in the confirmatory retesting of some samples. In all cases, the adjusted HIV prevalence estimates from the two types of Bayesian latent class models should be interpreted with caution. The dichotomous latent class model is strongly dependent on prior information on test error rates and the continuous latent class model assumes that the log-transformed signal-to-cutoff values are normally distributed. This latter assumption may have been violated in some surveys such as Sierra Leone and Uganda.

Despite the imperfections of methods used across all surveys, including our latent class models, there is strong agreement for most countries between interval estimates (high and/or complete overlap) from the different models. Although this is reassuring, further work is needed to improve the Bayesian latent class models in order to increase confidence in the estimates. Finally, imperfect laboratory testing is only one of several factors that could affect HIV prevalence estimates such as random sampling error, and selection and non-response bias.

List of abbreviations

AIS: AIDS indicator surveys

CDC: The US Centers for Disease Control and Prevention

CI: Confidence interval

CrI: Credible interval

CT: Confirmatory testing

DBS: Dried blood spot

DHS: Demographic and Health Survey

DRC: Democratic Republic of Congo

EIA: Enzyme immunoassay

HIV: Human immunodeficiency virus

IQC: Internal quality control

NAT: Nucleic acid test

OD: Optical density

RDT: Rapid diagnostic test

S/CO: Signal-to-cutoff

SD: Standard deviation

Se: Sensitivity

Sp: Specificity

UNAIDS: The Joint United Nations Program on HIV/AIDS

WHO: World Health Organization

1. Introduction

Accurate measurements of HIV prevalence are the cornerstone of HIV prevention activities (UNAIDS/WHO 2005). Nationally representative HIV seroprevalence surveys enable control programs to devise appropriate programmatic responses, monitor trends, and measure the impact of interventions (UNAIDS/WHO 2005). Since 2001, *The DHS Program* has conducted 71 HIV seroprevalence surveys in 39 countries. These surveys have become an important data source for individual researchers, national control programs, and the *Joint United Nations Program on HIV/AIDS* (UNAIDS).

HIV testing in these large surveys is logistically challenging with blood samples analyzed in a central laboratory located in the host country. The countries used different serological assays, but the HIV testing strategy employed usually consisted of two enzyme-immunoassays (EIA) and a third tie-breaker assay, which was usually a blot assay. Although the laboratory procedures for HIV serology testing in these population-based surveys followed the 2005 recommendations from UNAIDS and the World Health Organization (WHO), recent concerns about overestimation of HIV prevalence prompted a revision of the recommended testing algorithm (UNAIDS/WHO 2015; Fishel and Garrett 2016).

In response to these concerns, *The DHS Program* recently reviewed the quality of EIA-based HIV serosurvey data from 20 Demographic and Health Surveys (DHS) and AIDS Indicator Surveys (AIS) conducted between 2010 and 2014 (Fishel and Garrett 2016). This review examined agreement between the two EIA and the distribution of their signal-to-cutoff (S/CO) values. The findings suggested that the proportion of discrepant EIA results was higher than the desired benchmark in 7 of the 20 surveys. In addition, the proportion of weakly reactive samples on the first EIA, which may be more likely to be false positives, was higher than 20% in 9 of the 20 surveys. These results suggest that the HIV status could have been misclassified for some samples, although the magnitude of this problem likely varies with the specific survey.

Although future surveys conducted by *The DHS Program* will employ the newly recommended testing strategy (UNAIDS/WHO 2015; WHO 2015a), understanding the impact of potential misclassification on past surveys' prevalence estimates is important both for epidemic monitoring and assessing the effectiveness of current and future HIV interventions. The re-testing of specimens collected from more than 20 serosurveys is not logistically possible and is also hindered by potentially suboptimal storage conditions. In such cases, using statistical models to incorporate the uncertainty associated with the imperfect sensitivity and specificity of the testing algorithm could enable a more detailed appraisal of the laboratory data.

This report will present analyses that explore the impact of the imperfect accuracy of EIA. First, we briefly discuss how population-based HIV seroprevalence surveys are conducted by *The DHS Program* (Sections 2.1 to 2.3). Second, we provide a description of the different assays and the potential sources of misclassification (Sections 2.4 to 2.5). Third, we describe the dichotomous and continuous Bayesian latent class model specifications used to adjust HIV prevalence estimates (Section 2.6). Fourth, we present results from the Bayesian latent class models and contrast them with those that assume no misclassification (Section 3). Finally, we discuss the relevance of these findings for HIV surveillance and highlight some limitations of our work (Section 4).

2. Methods

2.1. HIV testing for population-based surveys estimates

According to the 2005 UNAIDS/WHO testing recommendations, all samples reactive on a first test should be confirmed using a second test (UNAIDS/WHO 2005). If these two tests are discrepant, a third test should be used as a tie breaker. The testing strategies employed over that period for surveys conducted by *The DHS Program* was consistent with these recommendations. In comparison, the updated WHO guidelines endorsed the use of different testing algorithms in settings with low and high HIV prevalence. For populations where the expected HIV prevalence is low (<5%), all samples that test positive on the first two EIA should be confirmed with a third assay. The assays should be chosen so that the first EIA has the highest sensitivity and the second EIA the highest specificity (UNAIDS/WHO 2005; Fishel and Garrett 2016; Urassa et al. 1999).

The high prevalence setting (>5%) algorithm now recommends mandatory repeat testing if the sample produces discrepant results on the first two assays. If the results are still discrepant, the sample is tested on a third assay. The guidelines for interpreting the results of the third assay have also changed since the 2005 testing strategy for both low and high prevalence settings. If the third assay is reactive, it is not recommended to use this result as the final HIV status because it selects for false positive results (UNAIDS/WHO 2015; WHO 2015a). Because such false positives could lead to the overestimation of HIV prevalence in population-based serosurveys (Hakim et al. 2016; Parekh and McDougal 2005), the individual is categorized as inconclusive (WHO 2015a).

The US Centers for Disease Control and Prevention (CDC) have also updated their recommendations for HIV testing in a clinical setting. The HIV-1 Western blots and HIV-1 indirect immunofluorescence assays are no longer included in the testing algorithm. Instead, indeterminate results produced from the EIA are tested with an HIV-1 nucleic acid test (NAT), which detects HIV RNA and not HIV antibodies (CDC/APHL 2014). Evidence now suggests that relying on Western blots or immunofluorescence assays for confirmation of positive samples on EIA can produce indeterminate results for early HIV infections (CDC/APHL 2014). The HIV NAT results reduce the number of indeterminate results because they distinguish between false positive EIA results and acute HIV infections. However, HIV NAT is not generally used for population-based HIV surveillance estimates (UNAIDS/WHO 2015).

2.2. Surveys conducted by *The DHS Program* and DHS testing algorithms

Detailed laboratory results from 23 surveys were provided by *The DHS Program*. All surveys except two (Dominican Republic and Haiti) were conducted in sub-Saharan Africa and one country (Rwanda) had two surveys conducted during that period. The sample size and HIV prevalence estimate for each survey are presented in Table 1.

To conduct the proposed analyses, we obtained the laboratory data files from each survey. The individual assay results of all samples from each survey were individually merged with CSPro to their DHS/AIS HIV recode file (Fishel and Garrett 2016). The laboratory files included individual results from all assays that were conducted during the surveys: the lab ID, plate ID, plate cutoff value, optical density (OD), and S/CO value for each sample, as well as basic demographical information for each participant.

Table 1. **Unweighted HIV prevalence estimates**

The unweighted HIV prevalence among men and women (age ≥15 years, upper age limit varies by survey), the unweighted total number of HIV positive participants, and the unweighted total number of participants who were tested for HIV, by survey

	Unweighted HIV prevalence among women and men age ≥15 years (upper age limit varies by survey❖)			
Survey	**Point estimate***	**95% confidence interval**	**Total number of HIV positive (unweighted)**	**Total number tested (unweighted)**
Sub Saharan Africa				
Burundi DHS 2010	1.9	(1.6-2.2%)	164	8588
Cameroon DHS 2011	4.6	(4.2-4.9%)	649	14,202
Chad DHS 2014-15	1.5	(1.3-1.8%)	166	10,726
Côte d'Ivoire DHS 2011-12	3.7	(3.3-4.1%)	336	9008
DRC DHS 2013-14	1.0	(0.9-1.2%)	177	17,638
Gabon DHS 2012	4.4	(4.0-4.8%)	483	10,992
Gambia DHS 2013	1.8	(1.5-2.1%)	136	7769
Ghana DHS 2014	1.9	(1.6-2.2%)	164	8848
Guinea DHS 2012	2.0	(1.7-2.3%)	164	8380
Malawi DHS 2010	10.2	(9.7-10.7%)	1420	13,910
Mali DHS 2012-13	1.1	(0.9-1.3%)	97	8861
Namibia DHS 2013	13.9	(13.2-14.7%)	1233	8858
Niger DHS 2012	0.5	(0.4-0.7%)	44	8628
Rwanda DHS 2010	3.2	(2.9-3.5%)	420	13,248
Rwanda DHS 2014-15	3.2	(2.9-3.5%)	415	12,940
Senegal DHS 2010-11	0.9	(0.8-1.1%)	93	9917
Sierra Leone DHS 2013	1.5	(1.3-1.7%)	222	14,600
Tanzania AIS/MIS 2011-12	4.6	(4.3-4.9%)	815	17,745
Togo DHS 2013-14	2.2	(1.9-2.5%)	199	9172
Uganda AIS 2011	7.0	(6.7-7.3%)	1495	21,367
Zambia DHS 2013-2014	13.4	(13.1-13.8%)	3901	29,006
Latin America/Caribbean				
Dominican Republic DHS 2013	1.0	(0.9-1.2%)	193	18,614
Haiti DHS 2012	2.3	(2.1-2.5%)	419	18,531

Note: The prevalence estimates in this table are unweighted survey results, and thus do not match the official survey prevalence estimates published in the DHS survey reports.

*The indeterminate results were not used in the HIV prevalence calculation except for Côte d'Ivoire and Guinea, where they were treated as negative. These results were included in the total number of individuals who were tested.

❖In most surveys, the ages for women ranged from 15-49 and men from 15-59. However, there are a few surveys that did not follow these age categories: Malawi (men age 15-54), Namibia (women and men age 15-64), Tanzania (women and men age 15-49), and Uganda (women and men age 15-59). All women and men in each country are included in the calculation of the HIV prevalence and in the following analysis, despite the varying age ranges.

The HIV assays and specific testing strategies used in these surveys vary by survey, and are selected in collaboration with the host country. The general testing strategy includes a combination of two EIA and a third tie-breaker assay for samples that are discrepant on the first two EIA – which is usually a Western Blot or a line immunoassay. It is also customary to randomly retest a certain proportion (5-10%) of samples that are negative on the first EIA as part of internal quality control (IQC). *The DHS Program* reviews laboratory procedures if greater than 1% of the IQC-selected samples test positive on the second EIA.

Testing also differs among surveys according to the HIV assays used and if repeat parallel testing was conducted when the two EIA results were discrepant. The description of the different assays and the characteristics of the testing algorithms are shown for each survey in Table 2. These surveys can be classified into four general categories. In the first general testing strategy (Type 1), all samples are tested with the first assay, and those found to be reactive are tested with a second assay. Both assays are EIA. Discrepant results on the first two assays are resolved with a third and definitive assay, which is either a Western Blot or line immunoassay. Finally, a certain proportion of samples that were initially non-reactive on the first assay are re-tested with the second assay as part of IQC (Figure 1). The second testing strategy (Type 2) is similar to the first, except that samples with discrepant results on the first two EIA are re-tested in parallel. Samples that are discrepant again are tested on a third, definitive assay (Figure 2). With the third testing strategy (Type 3), no third assay is involved, but the initially discrepant samples are tested again in parallel (Figure 3). With the fourth testing strategy (Type 4), there was no IQC and no repeat testing in parallel for initially discrepant results, and a third EIA served as the third, definitive assay instead of the usual Western Blot or line immunoassay (Figure 4).

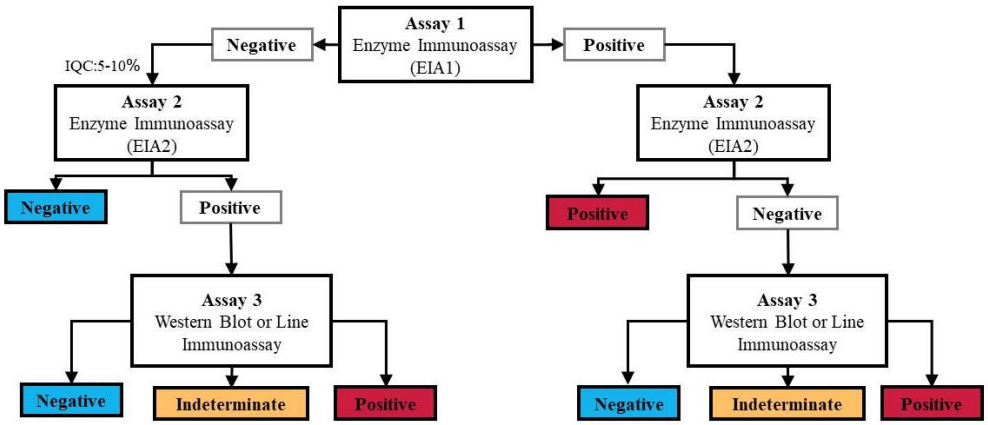

Figure 1. **General DHS testing strategy - Type 1.** This testing strategy was used in the following surveys: Burundi (DHS 2010)*, Cameroon (DHS 2011), Gabon (DHS 2012), Gambia (DHS 2013), Haiti (DHS 2012), Mali (DHS 2012-13), Namibia (DHS 2013*), Niger (DHS 2012)*, Rwanda (DHS 2010, 2014-15), Senegal (DHS 2010-11)*, and Sierra Leone (DHS 2013)*. (*A line immunoassay was used as the third assay instead of a Western Blot.)

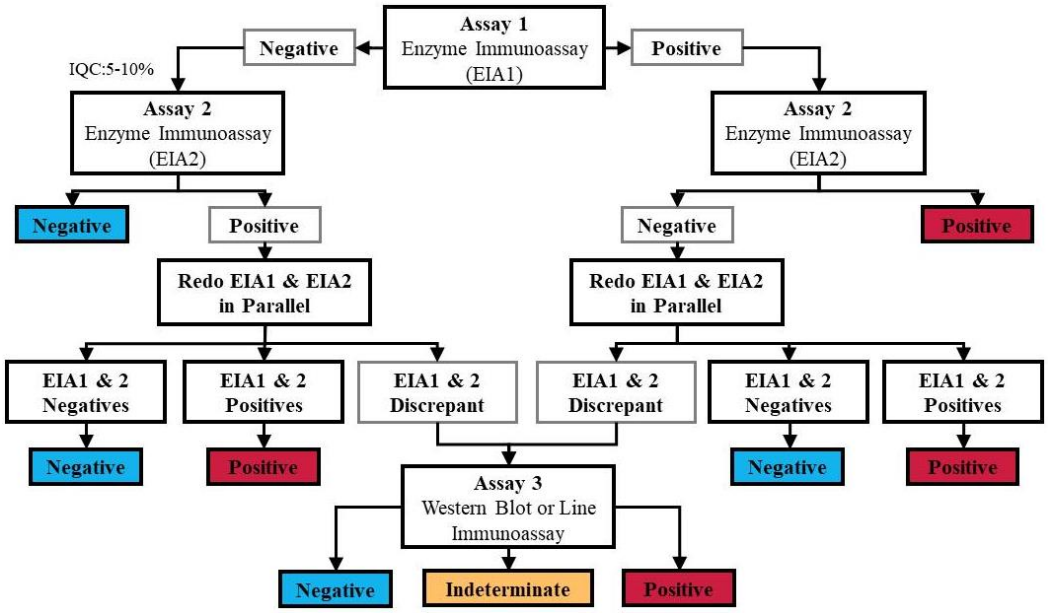

Figure 2. **General DHS testing strategy - Type 2.** This testing strategy was used in the following surveys: Chad (DHS 2014-15), Democratic Republic of the Congo (DHS 2013-14), Dominican Republic (DHS 2013), Ghana (DHS 2014)*, Malawi (DHS 2010), Tanzania (AIS/MIS 2011-12), Togo (DHS 2013-14)* and Zambia (DHS 2013-14). (*A line immunoassay was used as the third assay instead of a Western Blot.)

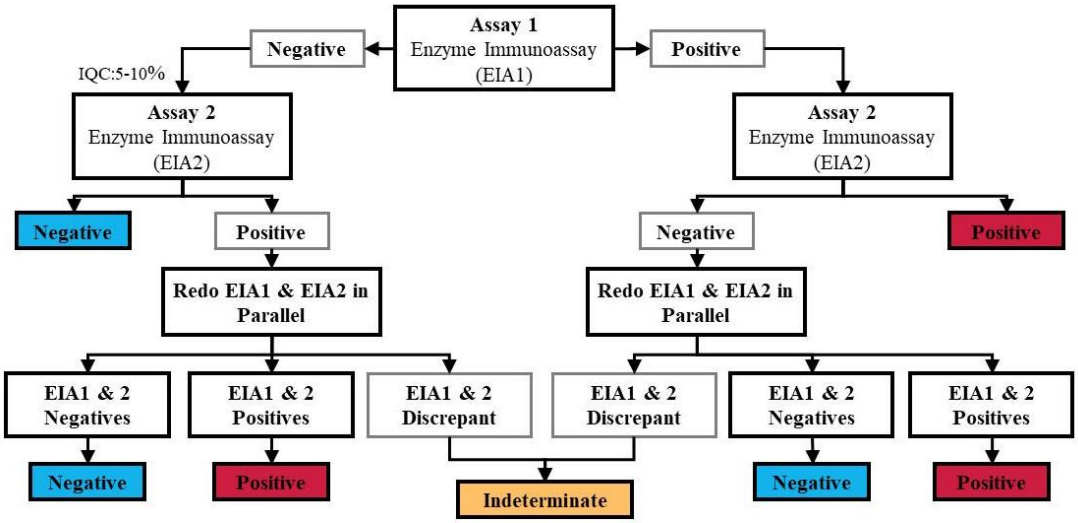

Figure 3. **General DHS testing strategy - Type 3.** This testing strategy was used in the following surveys: Côte d'Ivoire (DHS 2011-12) and Guinea (DHS 2012).

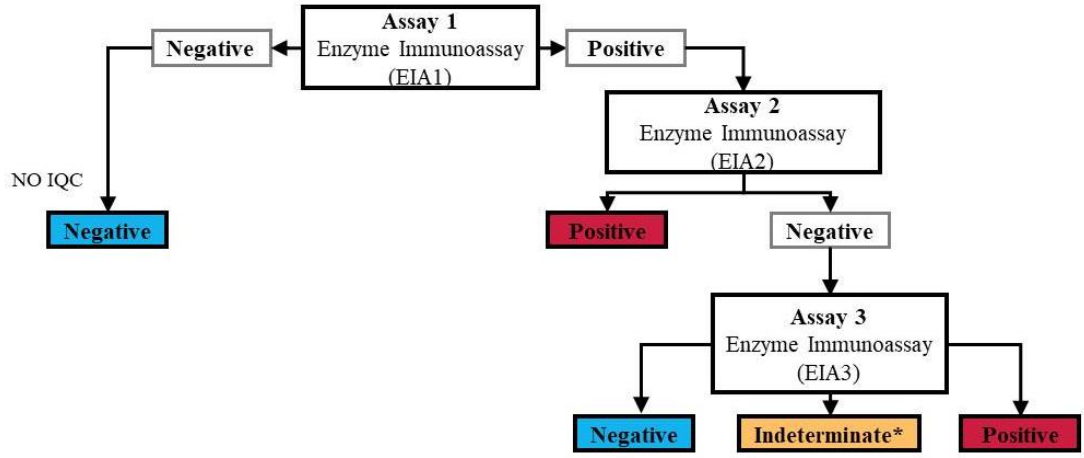

Figure 4. **General DHS testing strategy - Type 4.** This testing strategy was used in the following survey: Uganda (AIS 2011). (*One indeterminate result. Assay 3 was not performed on the sample.)

Table 2. DHS testing algorithms and HIV assays used

The name, manufacturer, and generation of the assays used in the HIV testing algorithms and specimen matrix, by survey

Survey	EIA 1	EIA 1 generation*	EIA 2	EIA 2 generation*	Parallel re-testing	Assay 3	Specimen matrix
Sub-Saharan Africa							
Burundi DHS 2010	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	No	Inno-Lia HIV I/II (Innogenetics)	DBS
Cameroon DHS 2011	Murex HIV Ag/Ab (DiaSorin)	4	Genscreen ULTRA HIV Ag/Ab (Bio-Rad)	4	No	New Lav Blot (Bio-Rad)	DBS
Chad DHS 2014-15	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	Yes	HIV 2.2 Blot (DiaSorin)	DBS
Côte d'Ivoire DHS 2011-12	Enzygnost Integral II (Siemens)	4	Vironostika HIV Ag/Ab (Biomérieux)	4	Yes	No third assay❖	DBS
DRC DHS 2013-14	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	Yes	HIV 2.2 Blot (DiaSorin)	DBS
Gabon DHS 2012	Genscreen Plus HIV Ag/Ab 5PL (Bio-Rad)	4	Enzygnost Integral II (Siemens)	4	No	HIV 2.2 Blot (DiaSorin)	DBS
Gambia DHS 2013	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	No	HIV 2.2 Blot (DiaSorin)	DBS
Ghana DHS 2014	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	Yes	Inno-Lia HIV I/II (Innogenetics)	DBS
Guinea DHS 2012	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	Yes	No third assay❖	DBS
Malawi DHS 2010	Vironostika HIV Uni-Form II Plus O (Biomérieux)	3	Enzygnost Anti-HIV 1/2 Plus (Siemens)	3	Yes	Western Blot 2.2 (Abbott Labs)	DBS
Mali DHS 2012-13	Vironostika HIV Uni-Form II Plus O (Biomérieux)	3	Enzygnost Integral II (Siemens)	4	No	HIV 2.2 Blot (DiaSorin)	DBS
Namibia DHS 2013	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	No	Inno-Lia HIV I/II (Innogenetics)	DBS
Niger DHS 2012	Vironostika HIV Uni-Form II Plus O (Biomérieux)	3	Enzygnost Anti-HIV 1/2 Plus (Siemens)	3	No	Inno-Lia HIV I/II (Innogenetics)	DBS
Rwanda DHS 2010	Vironostika HIV Ag/Ab (Biomérieux)	4	Murex HIV Ag/Ab (DiaSorin)	4	No	HIV 2.2 Blot (DiaSorin)	DBS
Rwanda DHS 2014-15	Vironostika HIV Ag/Ab (Biomérieux)	4	Murex HIV Ag/Ab combination (DiaSorin)	4	No	HIV 2.2 Blot (DiaSorin)	DBS
Senegal DHS 2010-11	Vironostika HIV Uni-Form II Plus O (Biomérieux)	3	Enzygnost Anti-HIV 1/2 Plus (Siemens)	3	No	Inno-Lia HIV I/II (Innogenetics)	DBS
Sierra Leone DHS 2013	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	No	Inno-Lia HIV I/II (Innogenetics)	DBS
Tanzania AIS/MIS 2011-12	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	Yes	HIV 2.2 Blot (DiaSorin)	DBS
Togo DHS 2013-14	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	Yes	Inno-Lia HIV I/II (Innogenetics)	DBS

Uganda AIS 2011	Murex HIV 1.2.0 (Abbott Labs)	3	Vironostika HIV Uni-Form II Plus O (Biomérieux)	3	No	Ani Labsystems HIVEIA	Plasma☆
Zambia DHS 2013-2014	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	Yes	Western Blot 2.2 (Abbott Labs)	DBS
Latin America/ Caribbean							
Dominican Republic DHS 2013	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	Yes	HIV 2.2 Blot (DiaSorin)	DBS
Haiti DHS 2012	Vironostika HIV Ag/Ab (Biomérieux)	4	Enzygnost Integral II (Siemens)	4	No	HIV 2.2 Blot (DiaSorin)	DBS

DBS= Dried blood spot; EIA=Enzyme immunoassay.

Some microplate readers have an upper detection limit and all optical densities (OD) above this limit are hence censored (usually 3 or 4). The surveys conducted in the Democratic Republic of the Congo, Ghana, Guinea, Namibia, Rwanda 2014-15, Dominican Republic, and Haiti all had an upper OD value censored at 3. Other surveys include Malawi (OD values censored at 3 and 4), Rwanda 2010 (OD values censored at 4) and Zambia (there are 9 OD values for the first EIA that are larger than 3).

*4th generation assays can detect the p24 antigen that is characteristically expressed shortly following HIV infections. These 4th generation assays have a shorter test window period of detection than 3rd generation assay.

❖Côte d'Ivoire (DHS 2011-12) and Guinea (DHS 2012) did not have a third assay. If the specimen had discrepant results on assay 1 and 2, it was considered negative when calculating HIV prevalence.

☆ Specimen matrix was plasma instead of DBS. If respondents refused venipuncture, they could provide a DBS specimen from a finger prick. HIV tested specimens were therefore a mix of plasma and DBS. Fewer than 2% of specimens were DBS.

2.3. DHS testing standards

In all surveys conducted with the assistance of *The DHS Program*, the laboratory analyses are conducted in the host country. The laboratory facility is selected by a DHS biomarker specialist and local agencies to ensure that the chosen laboratory can conduct high-quality testing and can handle a large volume of specimens. A laboratory training manual was developed by *The DHS Program* to assist host countries in conducting their population-based seroprevalence surveys. In all countries, *The DHS Program* invests heavily in training the laboratory technicians and by providing supplies and equipment for the laboratory in order to enhance the countries' capacity for quality testing.

Except for the 2011 AIS in Uganda, the type of sample collected from each individual was a dried blood spot (DBS), in which a small amount of blood is collected from a finger prick on filter paper. Hakim et al. recently suggested that false positivity could be magnified with DBS (Hakim et al. 2016). In Uganda, field staff collected plasma samples by venipuncture. However, participants could refuse and provide a DBS specimen instead. Field personnel were also trained on DBS collection. Further information on specimen collection can be found in *The DHS Program* standard biomarker training manual (MEASURE DHS 2012).

2.4. Description of the assays used in the DHS testing algorithms

The 23 DHS surveys used nine EIA and three supplementary assays (one line immunoassay and two Western Blots), as described below.

Enzyme immunoassays

The EIA for HIV are used for the qualitative detection of HIV antigens or antibodies against HIV in biological samples. The most common types of EIA use the indirect and sandwich methods. EIA for HIV detection utilize the sandwich technique for the detection of HIV antigens and/or antibodies elicited from an HIV infection. The EIA used by *The DHS Program* are sandwich assays, in which the antigen of interest, or capture antibody, is coated on the bottom of the wells of a microtiter plate and the sample is then added to the well where the antigens and/or antibodies bind. A conjugate containing polyclonal antibodies against the bound antibody and/or antigen detects the presence of HIV-specific antibodies or antigens bound to the wells of the microtiter plate (Truant 2016).

Testing with EIA begins with the transportation of filter papers to the laboratories, where blood is then eluted from the individual DBS sample and the eluate is prepared for analysis (Grüner, Stambouli, and Ross 2015). The eluate from each DBS sample that represents a different individual is then added to a corresponding well in the microtiter plate to allow the antibodies and/or antigens in the sample to bind to the bottom of the wells of microtiter plate. Any unbound antibodies and/or antigens are removed by washing, and the secondary antibody, which is attached to the detection enzyme, is added to the plate. In most of the following assays, horseradish peroxidase (HRP) is used as the enzyme label. After the excess antibody is removed, the reaction between HIV proteins in the sample and the plate is observed by using the chromogen tetramethylbenzidine dihydrochloride (TMB) which reacts with the enzyme to produce a blue to yellow color. The color produced by the enzymatic reaction between the enzyme and TMB is measured at 450 nm with a reference wavelength of 630 nm using a microplate reader. Some microplate readers have an upper detection limit, and all optical densities (OD) values above this limit are censored. The color intensity is approximately proportional to the concentration of HIV antigen and/or HIV antibodies in the test sample. For the assay run to be valid, the control sera that are tested on the microtiter plate along with the samples must comply with the specifications provided by the manufacturer. That is, the OD of the positive and negative controls must meet certain criteria to ensure that the test is working correctly. The mean OD of the control samples can then be used to calculate the assay's cutoff value, as per the manufacturer's guidelines (Table 3). This cutoff value is then used to qualitatively define a sample as non-reactive (negative) or reactive (positive).

Positive EIA samples can also be categorized by their reactivity depending on their S/CO values. A *highly reactive* specimen usually has a S/CO value greater than 5, while a S/CO value between 3 and 5 can be considered *moderately reactive* and a S/CO value of 1 to 3 can be considered *weakly reactive* (Nkengasong et al. 1999). Although all three categories are considered positive, specimens with a S/CO value less than 5.0 were found to be more likely to be false positives in the literature (Fanmi et al. 2013; Nkengasong et al. 1999; Urassa et al. 1999). Despite the relationship between the strength of this EIA reaction and the true HIV status of a sample, the S/CO value cannot differentiate between true and false positive results. False positive samples with high S/CO values have been documented in the literature as well as true positive samples that have reacted weakly on an EIA (Chacon, Mateos, and Holguin 2017; Fanmi et al. 2013; Hakim et al. 2016; Zhang et al. 2013).

All nine EIA used in the 23 surveys are listed in Table 3 with their respective characteristics. All third and fourth generation assays can detect antibodies against HIV-1, HIV-2 and HIV-1 subtype O antigens in human serum or plasma. Only fourth generation assays can detect the p24 HIV antigen (Table S1). The ability to detect the p24 antigen allows the fourth generation assay to identify more recent HIV infections because this antigen can be detected earlier in the seroconversion process.

Sensitivity and specificity of enzyme immunoassays

There is a range of sensitivities and specificities reported in the literature for each EIA. In general, EIA are highly sensitive, although test accuracy can vary with laboratory conditions (Constantine 2006; Peeling et al. 2010). The nine assays used in the 23 surveys have reported sensitivity and specificity estimates higher than 99.8% and 98.3%, respectively (Table 3). These estimates are usually reported by the assay's manufacturer, independent organizations such as WHO, or from scientific studies. In general, sensitivity and specificity estimates are derived under optimal laboratory conditions by trained lab technicians, and we can expect that lower test properties may result from less stringent laboratory conditions and/or less experienced users (see Section 2.5).

Table 3. Properties of HIV enzyme immunoassays

All enzyme immunoassays (EIA) used in the 23 surveys conducted by *The DHS Program* with their respective characteristics. The generation, sensitivity, specificity, cutoff calculation, and number/type of controls are specific to the assay itself, whereas the range of signal-to-cutoff value and the delta values are in relation to the distribution of optical densities of each test.

Assay name (manufacturer)	Sensitivity (95% CI)	Specificity (95% CI)	Cutoff calculation	Confirmed HIV (+) S/CO mean (range)	Delta values (δ)	Controls
Enzygnost Anti-HIV 1/2 Plus (Siemens/Dade Behring)*	100% (99.6-100%) (WHO 1999; Beelaert et al. 2002) 99.75% (98.60-100%) (Weber et al. 2002) 100% (99.44-100%) (Nkengasong et al. 1999)	99.7% (99.1-100%) (WHO 1999; Beelaert et al. 2002) 99.94% (99.80-99.99%) (Weber et al. 2002) 99.75% (98.64-99.99%) (Nkengasong et al. 1999)	NC+0.4 (WHO 1999)	7.5 (2.3-18.5) (Urassa et al. 1999) 8.0 (2.3-8.6) (Aboud et al. 2006)	δ^+: 19.1; δ^-: -6.6 (WHO 1999; Beelaert et al. 2002) δ^+: 46.1; δ^-: -5.8 (Nkengasong et al. 1999) δ^+: 5.0; δ^-: -7.7 (Zeh et al. 2011)	N: 4, P: 2, IC: No (WHO 1999)
Enzygnost Integral II (Siemens/Dade Behring)	100% (97.7-100%) (WHO 2004)	100% (98.8-100%) (WHO 2004)	PC*0.23 (WHO 2004)	13.7 (5.3-31.0) (Urio et al. 2015)	N/A	N: 2, P: 2, IC: Yes (WHO 2004)
Genscreen Plus HIV Ag/Ab (Bio-Rad)	100% (97.7-100%) (WHO 2004) 100% (98.8–100%) (Piwowar-Manning et al. 2015)	98.3% (96.1–99.4%) (WHO 2004) 96.7% (94.0–98.4%) (Piwowar-Manning et al. 2015)	NC+0.2 (Bio-RadLaboratories 2011)	N/A	N/A	N: 1, WP: 3, P: 1, IC: Yes (WHO 2004)
Genscreen ULTRA HIV Ag/Ab (Bio-Rad)	100% (99.2-100%) (WHO 2013) 100% (98.3-100%) (Francis 2013)	99.24% (98.2-99.8%) (WHO 2013) 99.90% (99.75-99.96%) (Francis 2013)	NC+0.2 (WHO 2013)	N/A	δ^+: 12.9; δ^-:-4.7 (Francis 2013)	N: 3, P: 2 (WHO 2013)
HIV EIA (Ani Labsystems)*	100% (99.6-100%) (WHO 1998; Beelaert et al. 2002)	99.4% (98.6-100%) (WHO 1998; Beelaert et al. 2002)	0.3(PC-BC) + BC (WHO 1998)	N/A	δ^+: 14.2; δ^-: -3.9 (WHO 1998; Beelaert et al. 2002)	N: 1, P: 5, B: 2, IC: No (WHO 1998)
Murex HIV 1.2.O (Abbott Labs/ DiaSorin)*	100% (99.97-100%) (Abbott Laboratories 2007) 100% (99.44-100%) (Nkengasong et al. 1999)	99.91% (99.82-99.96%) (Abbott Laboratories 2007) 100% (99.10-100%) (Nkengasong et al. 1999)	NC+0.2 (Abbott Laboratories 2007)	N/A	δ^+: 22.8; δ^-:-3.0 (Nkengasong et al. 1999) δ^+: 4.5; δ^-: -6.3 (Zeh et al. 2011)	N: 3, P: 2 (Abbott Laboratories 2007)
Murex HIV Ag/Ab (DiaSorin)	100% (99.2-100%) (WHO 2015c) 100% (98.83-100%) (Lemee et al. 2014)	99.7% (98.9-100%) (WHO 2015c) 98.9% (97.5-99.6%) (Lemee et al. 2014)	NC+0.15 (WHO 2015c)	12.2 (5.2-16.0) (Urio et al. 2015) 13.3 (5.3-18.6) (Aboud et al. 2006)	N/A	N: 3, P: 1-2 (WHO 2015c)
Vironostika HIV Ag/Ab (Biomérieux/ Organon Teknika)	100% (97.7-100%) (WHO 2004)	99.0% (97.1-99.8%) (WHO 2004)	NC+0.10 (WHO 2004)	15.5 (12.1-19.0) (Urio et al. 2015) 15.8 (10.8-20.7) (Aboud et al. 2006)	N/A	N: 3, P: 1, IC: Yes (WHO 2004)
Vironostika HIV Uni-Form 2 Plus O (Biomérieux/ Organon Teknika)*	100% (99.6-100%) (WHO 1999; Beelaert et al. 2002) 100% (99.44-100%) (Nkengasong et al. 1999)	100% (99.7-100%) (WHO 1999; Beelaert et al. 2002) 99.50% (98.24-99.94%) (Nkengasong et al. 1999)	NC+0.10 (WHO 1999)	14.8 (6.5-21.2) (Aboud et al. 2006)	δ^+: 17.2; δ^-: -4.1 (WHO 1999; Beelaert et al. 2002) δ^+: 12.6; δ^-: -4.4 (Nkengasong et al. 1999) δ^+: 5.3; δ^-: -5.1 (Zeh et al. 2011)	N: 3, P: 1-2, IC: No (WHO 1999)

NC: Mean optical density of negative controls; PC: Mean optical density of positive controls; BC: Mean optical density of blank controls.
N: Negative control; P: Positive control; WPC: Weak positive control; B: Blank control; IC: Internal control.
*Indicates that this is a third generation assay, and does not detect the p24 antigen.

Confirmatory assays

Two different types of supplemental assays were used to resolve discrepant results: Western Blot and line immunoassays. A Western Blot is a qualitative test that detects HIV proteins on a membrane after their subsequent transfer from a gel. The manufacturer provides guidelines on band appearance patterns that indicate when the sample is HIV positive or negative (or indeterminate) and sometimes can distinguish the HIV subtypes. The HIV specific proteins are first separated by their molecular weight using gel electrophoresis. The molecules are then transferred (blotted) onto a membrane and detected by antibodies. In contrast with Western Blots, a line immunoassay procedure involves coating the HIV specific proteins as discrete lines on a nylon strip, which avoids electrophoresis. The test sample is added to the membrane and then washed away. The alkaline phosphatase-labeled anti-human IgG antibodies are incubated on the membrane and then bind to the HIV antigen/antibody complex that was previously formed. A combination of enzymes, 5-bromo-4-chloro-3-indolyl-phosphate and nitro blue tetrazolium (BCIP/NBT) produces a dark brown color. The reaction is stopped by removing the developing solution from the strips. The appearance of specific colored bands demonstrates the presence of HIV antibodies in the test sample.

Inno-Lia HIV I/II Score Line, New Lav Blot and HIV 2.2 Blot/Western Blot 2.2 detect the presence of antibodies against HIV-1, HIV-2, and HIV-1 subgroup O antigens in a human sera or plasma sample (Table 4). New Lav Blot has separate nitrocellulose strips for HIV-1 and HIV-2 proteins, although the HIV Blot 2.2 nitrocellulose strips contain all the HIV-1 proteins and a specific HIV-2 peptide (Table S2).

Sensitivity and specificity of confirmatory assays

Confirmatory assays are used because the lower specificity of the EIA can result in false positives. To fill this role, the assays must have very high specificity. Estimates of specificity reported in the literature are 100% for the three types of confirmatory assays used in the surveys, if the indeterminate results are excluded from the numerator and denominator of the specificity estimates (Table 4). Here again, the actual specificity of the tests can vary depending on laboratory conditions (Constantine 2006; Peeling et al. 2010).

Table 4. **Properties of supplemental assays**

Supplemental assays used in the 23 surveys conducted by *The DHS Program* with their respective characteristics: assay type, sensitivity and specificity. The results from these assays are qualitative.

Assay name (manufacturer)	Assay type	Sensitivity (95% CI)	Specificity (95% CI)
Inno-Lia HIV I/II Score Line (Innogenetics)	Line immunoassay	100% (98.6-100%) (WHO 2015b)	100% (98.0-100%) (WHO 2015b)
New Lav Blot (Bio-Rad/Sanofi Diagnostics Pasteur)	Western Blot	100% (98.1-100.0%) (WHO 2015b)	100% (96.8-100.0%) (WHO 2015b)
HIV 2.2 Blot/Western Blot 2.2 (DiaSorin/MP Diagnostics/Abbott Labs)	Western Blot	HIV-1: 99.52% (97.36-99.99%) (*MP Diagnostics 2003*) HIV-2: 95.33% (89.43-98.47%) (*MP Diagnostics 2003*) HIV-1: 100%* HIV-2: 100%❖	HIV-1/2: 93.50% (89.14-96.49%) (*MP Diagnostics 2003*) HIV-1/2: 100%☆

*If the indeterminate results are considered, then sensitivity (99.52%) remains the same where # of true positives is 208/209. If indeterminate results are disregarded then sensitivity becomes 208/208 (100%).

❖If the indeterminate results are considered, then sensitivity (95.33%) remains the same where # of true positives is 102/107. If indeterminate results are disregarded then sensitivity becomes 102/102 (100%).

☆If the indeterminate results are considered, then specificity (93.5%) remains the same where # of true positives is 187/200. If indeterminate results are disregarded then specificity becomes 187/187 (100%).

2.5. Potential sources of misclassification

Misclassification of HIV status can result from different factors such as limitations of the assay, operator-related factors, facility-related factors, suboptimal testing algorithms, and inadequate training (Johnson et al. 2015). Some of these errors may not be corrected in certain laboratories because of the lack of quality management systems or external quality assessments. Half of the studies that reported poor management and supervision in a WHO review of HIV testing services also reported poor management of supplies (Johnson et al. 2015). It is also possible that an assay is not suitable for detecting a circulating HIV variant in a particular region. In addition, cross-reactivity has been documented for antibodies from intercurrent infections, HIV subtypes, or even environmental exposure to the assay components (Johnson et al. 2015). Operator-related factors could include data reporting problems, incorrectly performing the assay, and misinterpreting the results. A common challenge is the interpretation of faint lines and weak reactive specimens on the qualitative confirmatory test (Western Blot and line immunoassay), which has been well documented and addressed in the literature (Gray et al. 2007; Sacks et al. 2012). Although the equipment is tested prior to testing blood specimens, clerical and user errors are still possible (Fishel and Garrett 2016). Poor record keeping and labeling, incorrect use of supplies such as incorrect buffer volumes, and technicians who do not follow appropriate reading times could lead to misdiagnosis (Johnson et al. 2015). The traffic of samples in a facility may also lead to misclassification—a high volume of specimens can overload the staff and a lower volume may lead to poor staff proficiency (Johnson et al. 2015). Finally, though the use of DBS with EIA is common, most EIA manufacturers have not validated their tests on this specimen type, and the use of DBS with these assays may magnify false positivity (Hakim et al. 2016).

2.6. Model specifications

In the absence of a gold-standard to assess HIV prevalence for large-scale household-based surveys, and the impossibility of retrospectively re-testing samples, a Bayesian latent class analysis can be used to adjust HIV prevalence estimates for potential misclassification (Joseph, Gyorkos, and Coupal 1995). By explicitly modeling the misclassification process and using external information on the accuracy of the different assays and the potential correlation between assays, it is possible to explore the extent to which HIV prevalence could have been over or under-estimated.

Two different types of models can be used to adjust HIV prevalence estimates, depending on whether all the information of the continuous assay results are used, or whether these results are first dichotomized by using the assays' manufacturer cutoffs. In this report, we explore both model types. The first class of Bayesian latent class model we considered uses only dichotomous test results. A more comprehensive way to model the tests results is to include the continuous score information from the OD of each assay. This second class of models uses the continuous information from the OD and the dichotomous information from the third assay (from which no continuous information is available) to provide adjusted prevalence estimates. In this section, we first describe the dichotomous class of model, and then the specification of the continuous latent class model.

Dichotomous assay model

Four different dichotomous assay models were derived, depending on the general testing strategy they used. We briefly present how the equations for the dichotomous assay model were derived. In the Type 1 testing strategy, there are two EIA used, with discrepant results resolved by using a third tie-breaker assay (no repeat testing) (Figure 5). For example, the equation for the probability of a sample being negative on the first EIA, and not selected for IQC, is equal to the product of the complement of the proportion tested for IQC and sum of the probabilities of being a false negative ($\pi[1-Se_1]$) and a true negative ($[1-\pi]Sp_1$), where the probabilities of being a false negative and true negative depends on the true unobserved prevalence (π) and the first EIA's properties in terms of sensitivity (Se_1) and specificity (Sp_1). If a sample that was negative on the first EIA was selected to be part of IQC and was found to be negative on the second EIA,

the equation would be equal to the product of the proportion of samples tested for IQC and the sum of probabilities of false negatives on both EIA ($\pi[1\text{-}Se_1][1\text{-}Se_2]$) and true negatives ($[1\text{-}\pi]Sp_1Sp_2$). The same logic is applied to all possible combinations of assays to derive the equations for the expected probability distribution (Figure 5).

The section below presents the equations for all four testing strategies.

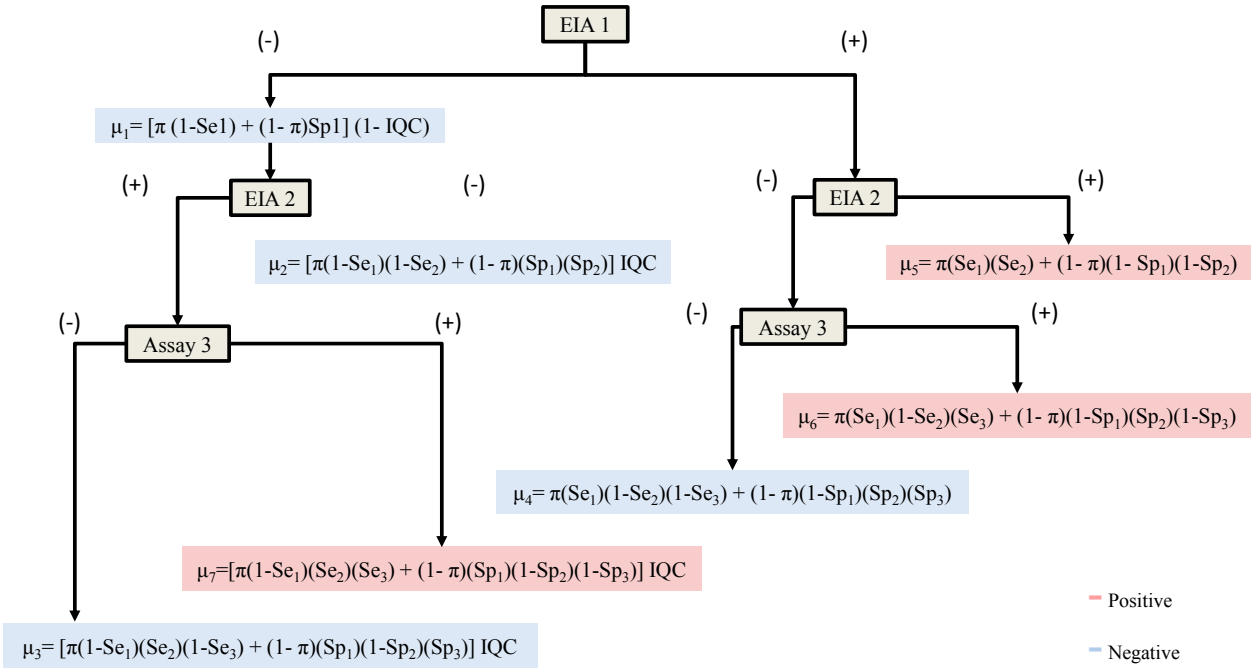

Figure 5. Flowchart describing the derivation of equations for the dichotomous Bayesian latent class model.

Type 1 testing strategy: two EIA, no repeat testing, with a third assay

The equations below describe the expected probabilities of positive and negative samples for each possible combination of assays in the Type 1 testing algorithm.

$$X \sim M(\mu, N)$$

$$
\begin{aligned}
EIA_1^- &: && \mu_1 = (1 - IQC)(\pi(1 - Se_1) + (1 - \pi)Sp_1) \\
EIA_1^- \ \&\ EIA_2^- &: && \mu_2 = IQC(\pi(1 - Se_1)(1 - Se_2) + (1 - \pi)Sp_1Sp_2) \\
EIA_1^- \ \&\ EIA_2^+ \ \&\ ASS_3^- &: && \mu_3 = IQC(\pi(1 - Se_1)Se_2(1 - Se_3) + (1 - \pi)Sp_1(1 - Sp_2)Sp_3) \\
EIA_1^+ \ \&\ EIA_2^- \ \&\ ASS_3^- &: && \mu_4 = \pi(Se_1)(1 - Se_2)(1 - Se_3) + (1 - \pi)(1 - Sp_1)Sp_2Sp_3 \\
EIA_1^+ \ \&\ EIA_2^+ &: && \mu_5 = \pi(Se_1Se_2) + (1 - \pi)(1 - Sp_1)(1 - Sp_2) \\
EIA_1^+ \ \&\ EIA_2^- \ \&\ ASS_3^+ &: && \mu_6 = \pi Se_1(1 - Se_2)Se_3 + (1 - \pi)(1 - Sp_1)Sp_2(1 - Sp_3) \\
EIA_1^- \ \&\ EIA_2^+ \ \&\ ASS_3^+ &: && \mu_7 = IQC\big(\pi(1 - Se_1)Se_2Se_3 + (1 - \pi)Sp_1(1 - Sp_2)(1 - Sp_3)\big)
\end{aligned}
$$

where X is a vector that contains the counts of samples for each possible combination of assays (Figure 5), as per the surveys' testing algorithm. These counts follow a multinomial distribution of size N and μ is a vector that contains the probabilities of observing these X counts; IQC is the proportion of samples negative on the first EIA (EIA_1) that are re-tested as part of the internal quality control with the second EIA (EIA_2); π is the true unobserved HIV prevalence; Se_1, Se_2, and Se_3 are the sensitivities of the first EIA, second EIA,

and third assays, respectively. Similarly, Sp_1, Sp_2, and Sp_3 are the specificities of the first EIA, second EIA, and the third assay, respectively.

For simplicity, we have assumed that the three assays are conditionally independent. In other words, we did not model the potential correlation between the three tests. A recent study evaluated the effect of conditional dependence on Bayesian latent class adjusted estimates of prevalence using dichotomous tests and found that the bias due to ignoring conditional dependence is generally small (Wang, Dendukuri, and Joseph 2017). In addition, we ran both correlated and non-correlated models for the continuous data (see below), and found the results to be similar.

Type 2 testing strategy: two EIA, repeat testing, with a third assay

The model for the Type 1 algorithm can be extended easily to account for repeat testing. Preliminary analyses (not shown) using an extended model accounting for repeat testing did not produce adjusted estimates that were very different from those estimated by the model above. Because of the assumption that the assays are conditionally independent, a model not including the repeated tests is more likely to produce good estimates, since we ignored potential conditional dependence. Hence, when using the Bayesian latent class model for dichotomous tests for the Type 2 testing strategy, we use the model above (Type 1) where the repeated parallel test results replace the initial test of the same type.

Type 3 testing strategy: two EIA, repeat testing, no third assay

$$
\begin{array}{ll}
 & X \sim M(\mu, N) \\
EIA_1^- : & \mu_1 = (1-IQC)(\pi(1-Se_1)+(1-\pi)Sp_1) \\
EIA_1^- \,\&\, EIA_2^- : & \mu_2 = IQC(\pi(1-Se_1)(1-Se_2)+(1-\pi)Sp_1Sp_2) \\
EIA_1^- \,\&\, EIA_2^+ : & \mu_3 = IQC(\pi(1-Se_1)Se_2+(1-\pi)Sp_1(1-Sp_2)) \\
EIA_1^+ \,\&\, EIA_2^+ : & \mu_4 = \pi(Se_1Se_2)+(1-\pi)(1-Sp_1)(1-Sp_2) \\
EIA_1^+ \,\&\, EIA_2^- : & \mu_5 = \pi Se_1(1-Se_2)+(1-\pi)(1-Sp_1)Sp_2
\end{array}
$$

Here again, the sample counts are modeled using a multinomial distribution. Tests that are still discrepant after repeat testing are considered indeterminate, and results from repeat testing are considered final.

Type 4 testing strategy: three EIA, no repeat testing, no internal quality control

$$
\begin{array}{ll}
 & X \sim M(\mu, N) \\
EIA_1^- : & \mu_1 = \pi(1-Se_1)+(1-\pi)Sp_1 \\
EIA_1^+ \,\&\, EIA_2^- \,\&\, EIA_3^- : & \mu_2 = \pi Se_1(1-Se_2)(1-Se_3)+(1-\pi)(1-Sp_1)Sp_2Sp_3 \\
EIA_1^+ \,\&\, EIA_2^+ : & \mu_4 = \pi(Se_1Se_2)+(1-\pi)(1-Sp_1)(1-Sp_2) \\
EIA_1^+ \,\&\, EIA_2^- \,\&\, EIA_3^+ : & \mu_5 = \pi Se_1(1-Se_2)Se_3+(1-\pi)(1-Sp_1)Sp_2(1-Sp_3)
\end{array}
$$

This final model, used only in Uganda, has the particularity that there was no internal quality control.

Prior distributions

We used a Bayesian framework to estimate HIV prevalence from the above models, which account for imperfect sensitivities and specificities of all tests regardless of the cutoff used for the serological classification. Bayesian models require the elicitation of prior distributions, which represent what is known about the unknown parameter values from expert opinion or past data. All our dichotomous models are non-identifiable because there are more parameters to estimate than degrees of freedom in the data. We chose to provide informative priors for the sensitivities and specificities of each of the three tests in order for the model to be essentially identifiable. This means that reasonable estimates can be produced for all parameters when combining the information in the data with the prior information. These priors were

chosen to account for the potential suboptimal laboratory conditions in which the samples could have been analyzed, and do not necessarily match those suggested by the manufacturers, which may reflect optimal lab conditions and highly experienced technicians. We chose a non-informative uniform prior for the true unknown HIV prevalence (equivalent to a *Beta*(1,1) density) and used informative beta distributions that match the 95% confidence intervals for sensitivity (95% confidence interval (95% CI: 98.0-99.9%)) and specificity (95% CI: 85.0-99.0%) of the EIA. For the third assay (Western Blot or line immunoassay), we assume that specificity would be slightly higher (95% CI: 99.0-99.9%). In this way, the combination of prior information on the tests with the testing data itself enables us to have identifiable models where the prior for the true unobserved HIV prevalence parameter has a very low information prior placed on it (i.e., uninformative).

Posterior distributions

According to Bayes' Theorem, we combined the likelihood function with the prior information to derive posterior densities, from which all inferences follow. Because the posterior distributions of the parameters of interest do not have closed form expressions, random samples from the posterior density are obtained using Markov Chain Monte Carlo (MCMC) simulations implemented with JAGS software (Hornik, Leisch, and Zeileis 2003; Plummer 2013). Three chains of 15,000 iterations were run after a burn-in of 7,000 iterations. These samples are then summarized to form inferences about all unknown parameters, including the prevalence. Parameters estimates are summarized with the median and 95% credible intervals (95% CrI; corresponding to the 2.5^{th} and 97.5^{th} percentiles) of their posterior distribution.

Continuous assay model

The main advantage to modeling continuous test results (the log[S/CO]) rather than dichotomizing at a cutoff value, as in the above analyses, is that continuous variables generally carry more information per observation as compared to discrete variables. Thus, sharper inferences on the estimated probabilities of each subject being truly positive or truly negative can lead to better prevalence estimates. To see why continuous values may carry more information, compare a weakly reactive sample, which should have a smaller chance of being a true positive to a sample that is strongly reactive, which should have a greater probability of being a true positive. Under dichotomous modeling, this detailed information is lost, which can lead to poorer estimates. Modeling continuous test results is more complex compared to using dichotomous test results. To facilitate understanding of the models, we introduce the following notation. Samples tested with one, two, three, and four (repeat testing in parallel) EIA are labeled S_1, S_2, S_3, and S_4, respectively. All discrepant samples on the first two EIA are tested with a third dichotomous assay (Western Blot or line immunoassay), except for the Uganda AIS 2011 survey, which used a third continuous EIA assay as a tie-breaker.

The OD from the continuous assays were modeled by first calculating their S/CO value. That is, the continuous OD were divided by their plate-specific cutoff value. Samples with a S/CO value greater than or equal to 1 were considered as reactive in the dichotomous analyses in the preceding section. By contrast, the continuous model specification does not explicitly use any threshold to categorize samples as non-reactive or reactive. Instead, the continuous model specification considers that each population of true positives and true negatives has their own densities of test results and that these two distributions overlap. Assuming that the distribution of positives lies to the right (higher values) of the distribution of negatives, very high S/CO values are more likely to have arisen from the positive density, while subjects with very low values are likely to be true negatives. However, subjects with intermediate values—values that lie in between the two density peaks—are of uncertain origin. For those values of uncertain origins, their probability of being positive will depend on their relative proximity to the positive and negative density peaks. These probabilities will be proportional to the heights of the density curves at the observed test value. These two densities can be modeled parametrically or non-parametrically. Nevertheless, because of the

large number of tests for each subject (up to four) and the potential for correlation between tests, we ruled out non-parametric models because no one has developed a model that accommodates four tests with correlations from a non-parametric viewpoint. In fact, it is also true that to date no one has developed a model that accommodates four tests with correlations from a parametric viewpoint, although this is a somewhat easier problem. For this report, we developed a novel model that accomplishes that. As is usual for these problems, we assumed normality of the densities conditional on true infection status. Since the S/CO values are not normally distributed, we log-transformed them (Feng et al. 2014; Qiu et al. 2008). Hence, we assumed that in sets S_2, S_3, and S_4, the vectors of test results X_i follow multivariate normal distributions of length 2, 3, or 4, respectively. While we realize that potential cross-contamination of samples, especially negative samples contaminated by positive residue, may lead to data that violate the normality assumption, we believe that our models will nevertheless produce reasonable estimates for most countries. Close inspection of all parameter estimates across analyses from the data from all countries led to a small number of unreasonable estimates that we believe were caused by imposing normality on contaminated samples. We attempted to resolve these problems by the judicious use of further prior information, for example, by restricting the range of means and/or standard deviations to plausible values when potential violations of multivariate normality (or surveys with censored OD values) led to highly implausible estimates.

The intuition behind our continuous latent class model is depicted in the figure below (Figure 6). Samples A, B, and C are tested with only one assay and each sample contribute information on the prevalence depending on their likelihood of arising from the distribution of true negatives (blue) or the distribution of true positives (red). As such, it does not directly use a single threshold value to differentiate positive from negative samples. In the example below, which depicts an ideal situation with near-perfect tests based on the distance between the positive and negative results, sample A would contribute a high probability of being part of the true negative distribution and a corresponding low probability of arising from the distribution of true positives. Sample B, which is almost exactly between the two distributions, and whose S/CO value would result in a positive classification with the dichotomous model despite its low reactivity, has a roughly equal chance of being a true negative or a true positive. It would therefore contribute the equivalent of "half a subject" to the prevalence (if the prior on the unobserved prevalence is uninformative). Finally, sample C is highly reactive and has a high probability of being a true positive and low probability of being a true negative. Thus, each subject's test score is classified not as a definitive positive or a negative result, but as a probability of being positive. The sum of these positive probabilities across all subjects determines the overall prevalence estimate.

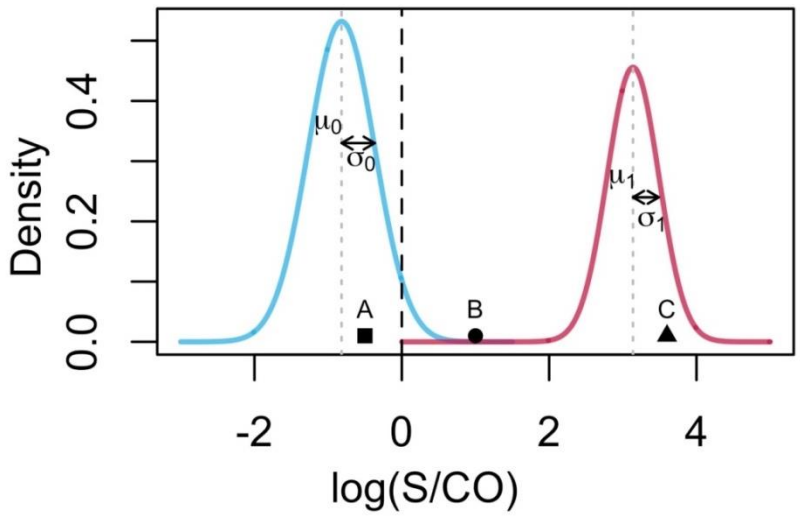

Figure 6. **Graphical representation of continuous model specification.** Graphical display of the continuous model specification where the signal-to-cutoff ratio on the logarithmic scale is displayed on the x-axis and the density of the normal distributions of the true negative (blue) with mean μ_0 and standard deviation σ_0 and of the true positive (red) with mean μ_1 and standard deviation σ_1. Points A, B, and C are hypothetical results that would lead to low, equivocal, and high probabilities of being positive, respectively, according to the ratios of the heights of the two densities at each point.

Likelihood for continuous assay results

Each sample is either truly positive or truly negative, which is modeled with the "*True Status*" latent variable T_i, a vector of length N, corresponding to the total number of samples in each respective survey. The contribution of T_i to the likelihood depends on the true unobserved prevalence π, as follows:

$$L_{Ti} = \pi^{Ti}(1-\pi)^{1-T_i}$$

The contribution of test scores (i.e., the log[S/CO]) from the i^{th} subject X_i to the likelihood is either $L_{i1|s} = L_1(x_i|\theta_s)$, $L_{i2|s} = L_2(x_i|\theta_s)$, $L_{i3|s} = L_3(x_i|\theta_s)$ or $L_{i4|s} = L_4(x_i|\theta_s)$, depending on the set $k=\{1,2,3,4\}$ to which sample i belongs and its latent true status $T_i=s$, where θ_s is the set of parameters associated with corresponding set and true status s. The likelihoods of these four different sets are presented below:

$$L_{i1|s} = \frac{1}{\sigma_1}\exp\left\{-\frac{(X_i-\mu_1)^2}{2\sigma_1^2}\right\}$$

$$L_{i2|s} = \frac{1}{|\Sigma_{2s}|}\exp\left\{-\frac{1}{2|\Sigma_{2s}|}(X_i-\mu)^\top\Lambda_{2s}(X_i-\mu)\right\} \text{ where } \Sigma_2 = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$$

$$L_{i3|s} = \frac{1}{|\Sigma_{3s}|}\exp\left\{-\frac{1}{2|\Sigma_{3s}|}(X_i-\mu)^\top\Lambda_{3s}(X_i-\mu)\right\} \text{ where } \Sigma_3 = \begin{pmatrix}\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3\\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3\\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2\end{pmatrix}$$

$$L_{i4|s} = \frac{1}{|\Sigma_{4s}|}\exp\left\{-\frac{1}{2|\Sigma_{4s}|}(X_i-\mu)^\top\Lambda_{4s}(X_i-\mu)\right\} \text{ where } \Sigma_4 = \begin{pmatrix}\sigma_1^2 & \rho_{11}\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{12}\sigma_1\sigma_2\\ \rho_{11}\sigma_1^2 & \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{12}\sigma_1\sigma_2\\ \rho_{12}\sigma_1\sigma_2 & \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{22}\sigma_2^2\\ \rho_{12}\sigma_1\sigma_2 & \rho_{12}\sigma_1\sigma_2 & \rho_{22}\sigma_2^2 & \sigma_2^2\end{pmatrix}$$

where μ is a vector of distribution means, whose value depends on the set k to which it belongs, and σ is the

standard deviation of that distribution; Λ_{ks} is the cofactor matrix corresponding to the covariance matrix Σ_{ks}, k={2, 3, 4}. These covariance matrices take into account potential conditional dependence between the different assays (the correlation parameter ρ). Since the models that account for conditional dependence gave similar results to the model that assumed independence, we chose to present HIV prevalence estimates from the latter model only, for simplicity. For these models, the covariance parameters on the off-diagonal entries of Σ_{ks} are simply replaced by zeros. The contribution of scores X_i to the likelihood, provided sample i pertain to set k, can be written as:

$$L_{X_i} = L_{ik|0}^{1-T_i} L_{ik|1'}^{T_i} \quad i \in S_k$$

Likelihood for confirmatory assay results

Each sample for which an additional confirmatory dichotomous assay was performed (Western Blot or line immune-assays) contributes an additional term to the likelihood. Given prior information on this confirmatory assay's sensitivity Se_W and specificity Sp_W, the contribution of this test (with binary results w_i, taking values of 0 if non-reactive and 1 if reactive) to the likelihood is given by:

$$L_{W_i} = \left[Se_W^{w_i}(1 - Se_W)^{1-w_i}\right]^{T_i} \left[(1 - Sp_w)^{w_i} Sp_W^{1-w_i}\right]^{(1-T_i)}$$

Prior distributions

As in the previous section, our Bayesian models require prior densities for each parameter. The prior distribution for the unobserved true HIV prevalence follows a uniform or, equivalently, a *Beta* distribution with:

$$\pi \sim Beta(1, 1)$$

The mean of the log(S/CO) for assay t among true positive samples (s=1) and true negative samples (s=0) are given normal prior distributions μ_{ts} with standard deviation σ_{ts} and where the mean of the normal distribution is ν_{ts}. As justified below, a minimum (δ_{ts}) and/or a maximum (ψ_{ts}) standard deviation was sometime used:

$$\mu_{ts} \sim N(\nu_{ts}, \sigma_{ts})$$
$$\nu_{ts} \sim N(\eta_{ts}, \xi_{ts})$$
$$f(\sigma_{ts}) = 1, and\ \delta_{ts} > \sigma_{ts} \leq \psi_{ts}$$

The different prior correlation terms (either between two tests or between the two measurements of a same test) are assumed to be uniform:

$$\rho \sim U(-1,1)$$

Finally, the sensitivity and specificity parameters of the confirmatory (dichotomous) assays are given truncated Beta distributions:

$$Se_W \sim Beta(1,1) * I(Lower_{Se}, 1)$$
$$Sp_W \sim Beta(1,1) * I(Lower_{Sp}, 1)$$

where $Lower_{Se}$ and $Lower_{Sp}$ represent prior lower limits, whose values varied between surveys and were typically 0.96 or 0.98. Preliminary analyses using non-truncated informative *Beta* priors on those parameters resulted in HIV prevalence estimates that were often much higher than the observed

(unadjusted) prevalence. Given the reported high sensitivity of EIA (UNAIDS/WHO 2015), it seemed unlikely that there would be such a high number of false negatives. For example, the posterior distribution for the sensitivity of the third confirmatory assays (Western Blot or line immunoassay) was as low as 79% (using a uniform prior), which was much lower than any reported estimates in the literature (Table 4). To better reflect the characteristics of the different assays, more informative prior distributions were elicited (Table 5). Specifically, truncated beta distributions were used where the lower limits of the distributions (indicator I) bounded the posterior distributions to high minimum values that are in line with their reported sensitivity and specificity.

We also provided a minimum value for the standard deviation (δ_{ts}) for the distribution of the true negative samples. This value ensured that the distribution of true negative would be wide enough to capture both samples that were non-reactive and those that were reactive and near the cutoff. For surveys that had censored OD values, we also found it necessary to place a further constraint on the maximum mean of the true negative density so that the S/CO ratio would not be higher than 44.7, the maximum S/CO value observed in the non-censored surveys. Without an upper limit, imputing values for these censored OD data points produced, in some countries, a mean that was artificially raised to unreasonable levels. For similar reasons, a maximum standard deviation (ψ_{ts}) for the distribution of true positives was also elicited for surveys with censored OD values. The values used for all priors are presented in Tables 5 and 6 and a full justification is provided in Text S1.

Posterior distributions

The special nature of the likelihood functions and priors limited the ability to fit our models in standard MCMC programs such as WinBUGS (Lunn et al. 2000), Nimble (Nimble 2017), or Stan (Carpenter et al. 2016). For example, Nimble does not allow random indices, and censoring can be problematic for programs such as WinBUGS. We therefore created our own MCMC routines in R to sample from the posterior distributions of the parameters of interest. This was achieved by generating random values for each parameter based on their respective full conditional posterior distribution. The full conditional distributions are presented in Text S2. For each analysis, we ran 1,000 burn-in iterations followed by 5,000 iterations for inference.

Table 5. **Priors for the true positive and true negative distributions of log-transformed signal-to-cutoff ratios for the nine enzyme immunoassays**

Assays	Prior distributions*
Enzygnost Integral II (Siemens)	$\mu_{t1} \sim N(v_{t1}, \sigma_{t1})$ $v_{t1} \sim N(log(13.7), 0.43)$, and $f(\sigma_{t1})=1$, and $\sigma_{t1} > 0$ $\mu_{t0} \sim N(v_{t0}, \sigma_{t0})$ $v_{t0} \sim N(log(0.31), 0.11)$, and $f(\sigma_{t0})=1$, and $\sigma_{t0} > 0.71$
Enzygnost Anti-HIV ½ Plus (Siemens)	$\mu_{t1} \sim N(v_{t1}, \sigma_{t1})$ $v_{t1} \sim N(log(7.7), 0.19)$, and, $f(\sigma_{t1}) = 1$, and $\sigma_{t1} > 0$ $\mu_{t0} \sim N(v_{t0}, \sigma_{t0})$ $v_{t0} \sim N(log(0.23), 0.43)$, and $f(\sigma_{t0})=1$, and $\sigma_{t0} > 0.89$
Genscreen Plus HIV Ag/Ab (Bio-Rad), Genscreen ULTRA HIV Ag/Ab (Bio-Rad), HIV EIA (Ani Labsystems), and Murex HIV 1.2.0 (Abbott Labs)	$\mu_{t1} \sim N(v_{t1}, \sigma_{t1})$ $v_{t1} \sim N(log(15), 0.47)$, and $f(\sigma_{t1}) = 1$, and $\sigma_{t1} > 0$ $\mu_{t0} \sim N(v_{t0}, \sigma_{t0})$ $v_{t0} \sim N(log(0.30), 0.35)$, and $f(\sigma_{t0})=1$, and $\sigma_{t0} > 0.73$
Murex HIV Ag/Ab Combination (DiaSorin)	$\mu_{t1} \sim N(v_{t1}, \sigma_{t1})$ $v_{t1} \sim N(log(12.7), 0.40)$, and $f(\sigma_{t1}) = 1$, and $\sigma_{t1} > 0$ $\mu_{t0} \sim N(v_{t0}, \sigma_{t0})$ $v_{t0} \sim N(log(0.30), 0.35)$, and $f(\sigma_{t0}) = 1$ and $\sigma_{t0} > 0.73$
Vironostika HIV Ag/Ab (Biomérieux)	$\mu_{t1} \sim N(v_{t1}, \sigma_{t1})$ $v_{t1} \sim N(log(21.5), 0.63)$, and $f(\sigma_{t1}) = 1$, and $\sigma_{t1} > 0$ $\mu_{t0} \sim N(v_{t0}, \sigma_{t0})$ $v_{t0} \sim N(log(0.33), 0.26)$, and $f(\sigma_{t0}) = 1$ and $\sigma_{t0} > 0.67$
Vironostika HIV Uni-Form II Plus O (Biomérieux)	$\mu_{t1} \sim N(v_{t1}, \sigma_{t1})$ $v_{t1} \sim N(log(19.4), 0.58)$, and $f(\sigma_{t1})=1$ and $\sigma_{t1} > 0$ $\mu_{t0} \sim N(v_{t0}, \sigma_{t0})$ $v_{t0} \sim N(log(0.33), 0.26)$, and $f(\sigma_{t0})=1$ and $\sigma_{t0} > 0.67$

*For surveys with censored optical density (OD) values, further constraints were placed on the maximum value that can take the mean and standard of the distribution of true positive samples. If OD values were censored, the mean of true positive (v_{t1}) could not be higher than log (44.7) – the maximum observed S/CO value - and the standard deviation (σ_{t1}) larger than 1.63 (on the log(S/CO) scale), the latter based on having less than 1% of the true positive distribution below the cutoff (i.e., a S/CO below 1).

Table 6. **Priors for the sensitivity and specificity of the dichotomous confirmatory assays**

Assays	Prior distributions
Inno-Lia HIV I/II Score Line Immunoassay (Innogenetics)	$Se_w \sim Beta(1,1)*I(0.986, 1)$ $Sp_w \sim Beta(1,1)*I(0.980, 1)$
New Lav Blot (Bio-Rad)	$Se_w \sim Beta(1,1)*I(0.981, 1)$ $Sp_w \sim Beta(1,1)*I(0.968, 1)$
HIV 2.2 Blot (DiaSorin) / Western Blot 2.2 (Abbott Labs)	$Se_w \sim Beta(1,1)*I(0.982, 1)$ $Sp_w \sim Beta(1,1)*I(0.980, 1)$

3. Results

To better understand agreement between the different assays, plots of the log-transformed S/CO values of the two EIA and confirmatory assay results were first produced (Figures 7A, 7B, 7C, and 7D). If the testing algorithm involved repeat testing, the final EIA values were plotted. These plots are divided into four quadrants. The bottom left quadrant includes all samples that were negative on the first EIA, that were selected for testing with a second EIA, and that were found to be non-reactive as part of the IQC procedure. The upper left quadrant has the samples that were initially non-reactive on the first EIA that were found reactive on the second EIA as part of the IQC procedure. The upper right quadrant includes all samples reactive on both EIA. Finally, the bottom right quadrant includes all samples that were reactive on the first EIA and non-reactive on the second EIA. All discrepant results (upper left and bottom right quadrants) were tested with a third confirmatory assay with the final results plotted with different colors (blue squares for negatives and red circles for positive samples). The medians of the posterior means, along with the 50% and 95% coverage intervals of the distributions (based on their median standard deviations) of the true negative and true positive distributions are also plotted on the margins of these graphs (Figures 7A, 7B, 7C, and 7D).

For many surveys, we see a clear clustering of almost all samples in the bottom left and top right quadrants. This indicates that samples were either non-reactive or strongly reactive on both EIA. For surveys with such clear clustering of positive and negative samples, we do not expect our misclassification adjustments to have much impact on HIV prevalence estimates. This group includes the following surveys: Côte d'Ivoire DHS 2011-12, Chad DHS 2014-15, DRC DHS 2013-14, Dominican Republic DHS 2013, Ghana DHS 2014, Guinea DHS 2012, Malawi DHS 2010, Mali DHS 2012-13, Rwanda DHS 2010, Rwanda DHS 2014-15, Tanzania AIS/MIS 2011-12, and Togo DHS 2013-14. An additional two surveys had relatively clear clustering of strongly reactive samples, although the agreement was slightly lower than for the preceding surveys: Gambia DHS 2013 and Senegal DHS 2010-11.

Surveys for which agreement was suboptimal and/or there was no clear clustering of strongly non-reactive or reactive samples include: Burundi DHS 2010, Cameroon DHS 2011, Gabon DHS 2012, Haiti DHS 2012, Namibia DHS 2013, Niger DHS 2012, Sierra Leone DHS 2013, Uganda AIS 2011, and Zambia DHS 2013-14. Of those, HIV testing results from the Uganda AIS 2011 and the Zambia DHS 2013-14 appear the most problematic. In Uganda, the 2011 AIS did not include any IQC and the third confirmatory assay was another EIA instead of the more specific Western Blot or line immunoassay. Further, a high number of samples have discordant results and there is no clear clustering of reactive samples. In Zambia, the 2013-14 DHS had higher agreement between the two EIA, although the high number of weakly reactive samples could suggest that a certain proportion of these are false positive.

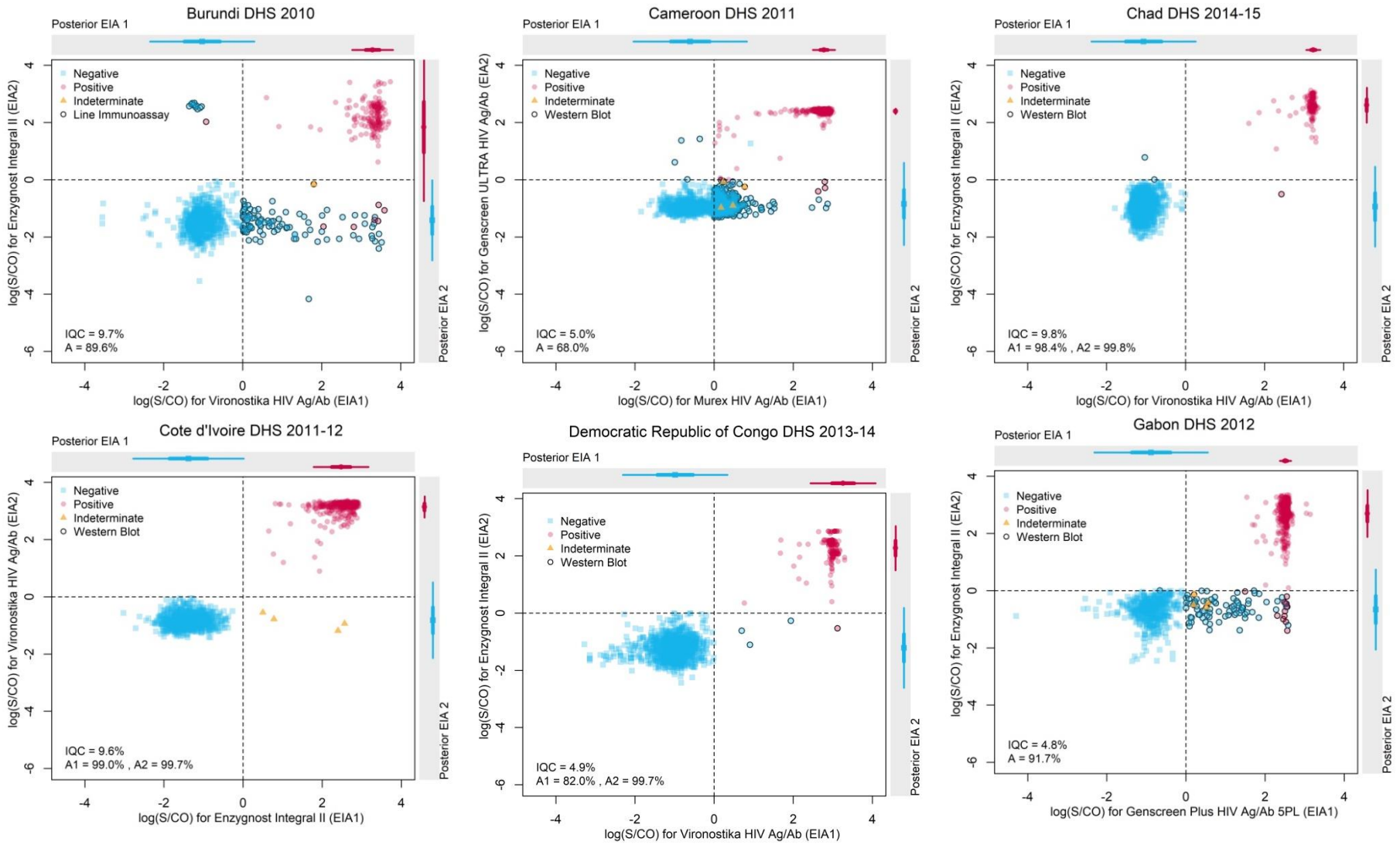

Figure 7A. Distribution of log transformed signal-to-cutoff ratio from samples tested on both enzyme immunoassays with results from the third assay.

Final HIV status is represented as the shape and color of each plot point. Samples tested on Western Blot or line immunoassays in the applicable surveys are circled in black. Côte d'Ivoire DHS 2011-12 did not have a third assay. A1/A2=Agreement between EIA 1 and EIA 2 before and after repeat testing respectively; A=Agreement for surveys without repeat testing; IQC=Proportion of negative samples on EIA1 selected for internal quality control. Negative and positive posterior distributions, from the continuous latent class model, are indicated on the top (EIA1) and right side (EIA2) of the graph (grey areas), where the point corresponds to the mean of the distribution, and the thick and thin lines represent 50% and 95% of the highest posterior density of the distributions.

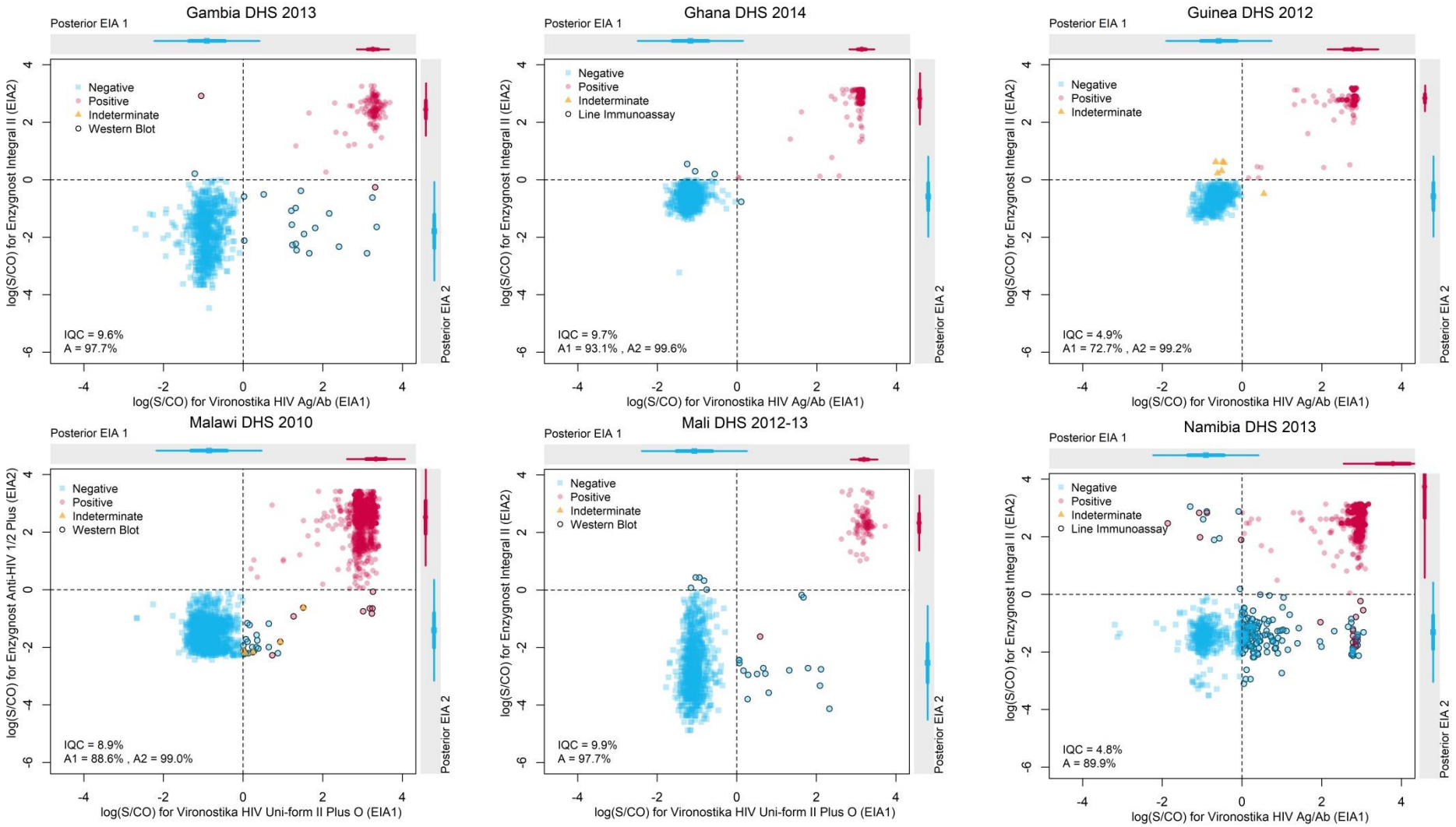

Figure 7B. Distribution of log transformed signal-to-cutoff ratio from samples tested on both enzyme immunoassays with results from the third assay.

Final HIV status is represented as the shape and color of each plot point. Samples tested on Western Blot or line immunoassays in the applicable surveys are circled in black. Guinea DHS 2012 did not have a third assay. A1/A2=Agreement between EIA 1 and EIA 2 before and after repeat testing respectively; A=Agreement for surveys without repeat testing; IQC=Proportion of negative samples on EIA1 selected for internal quality control. Negative and positive posterior distributions, from the continuous latent class model, are indicated on the top (EIA1) and right side (EIA2) of the graph (grey areas), where the point corresponds to the mean of the distribution, and the thick and thin lines represent 50% and 95% of the highest posterior density of the distributions.

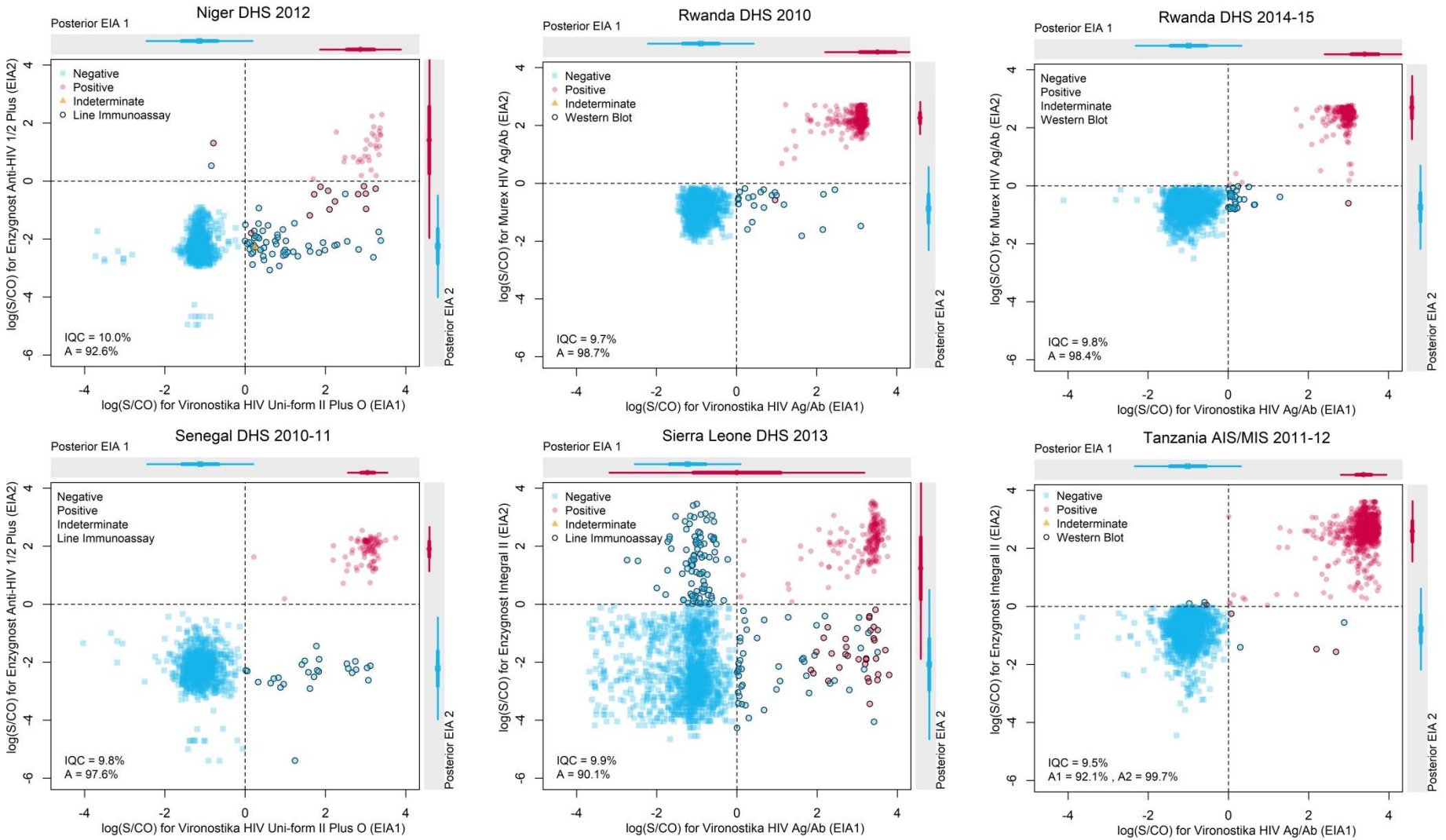

Figure 7C. Distribution of log transformed signal-to-cutoff ratio from samples tested on both enzyme immunoassays with results from the third assay.

Final HIV status is represented as the shape and color of each plot point. Samples tested on Western Blot or line immunoassays in the applicable surveys are circled in black. A1/A2=Agreement between EIA 1 and EIA 2 before and after repeat testing respectively; A=Agreement for surveys without repeat testing; IQC=Proportion of negative samples on EIA1 selected for internal quality control. Negative and positive posterior distributions, from the continuous latent class model, are indicated on the top (EIA1) and right side (EIA2) of the graph (grey areas), where the point corresponds to the mean of the distribution, and the thick and thin lines represent 50% and 95% of the highest posterior density of the distributions.

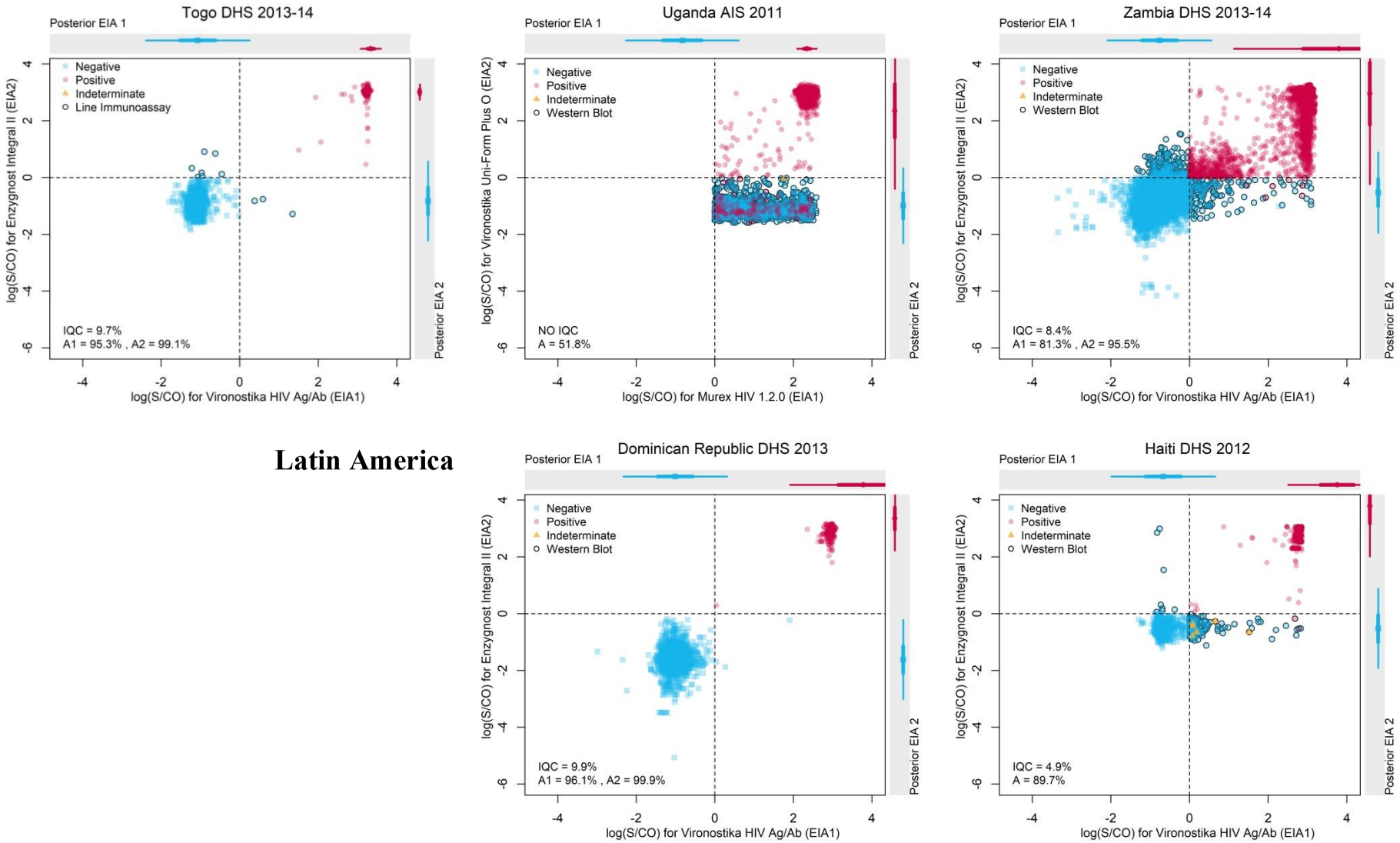

Figure 7D. Distribution of log transformed signal-to-cutoff ratio from samples tested on both enzyme immunoassays with results from the third assay.

Final HIV status is represented as the shape and color of each plot point. Samples tested on Western Blot or line immunoassays in the applicable surveys are circled in black. Dominican Republic DHS 2013 did not perform any Western blots for discrepant results after repeat testing. Uganda DHS 2011 did not have internal quality control (IQC) and had a third EIA instead of a Western blot. A1/A2=Agreement between EIA 1 and EIA 2 before and after repeat testing respectively; A=Agreement for surveys without repeat testing; IQC=Proportion of negative samples on EIA1 selected for internal quality control. Negative and positive posterior distributions, from the continuous latent class model, are indicated on the top (EIA1) and right side (EIA2) of the graph (grey areas), where the point corresponds to the mean of the distribution, and the thick and thin lines represent 50% and 95% of the highest posterior density of the distributions.

Prevalence estimates from the latent class models

Surveys with optimal assay concordance

For the category of surveys with optimal agreement between assays and clear clustering of non-reactive and strongly reactive samples, our Bayesian latent class adjustments – either using the dichotomous or the continuous model – did not have much effect on the prevalence estimates and all uncertainty intervals were overlapping (Figures 8A and 8B). For example, Côte d'Ivoire had an observed HIV prevalence of 3.7% (95% CI: 3.3-4.1%), an adjusted prevalence from the dichotomous model of 3.8% (95% CrI: 3.4-4.2%), and an adjusted prevalence from the continuous model of 3.7% (95% CrI: 3.3-4.1%) (Table 7).

In Tanzania, agreement between the two EIA was 99.7% after repeat testing (92.1% before repeat testing). There were a few weakly reactive samples on both EIA, however. The unadjusted and dichotomous model adjustments resulted in the same prevalence of 4.6%. This was expected because the dichotomous model does not qualitatively differentiate between weakly reactive and strongly reactive samples. The continuous model, however, gave a higher likelihood of being truly negative to those few weakly reactive samples, and resulted in a slightly lower HIV prevalence of 4.5%. Because the great majority of samples were strongly reactive on both EIA, however, the adjustment with the continuous model had little effect on the HIV prevalence estimate and all uncertainty intervals overlapped.

Surveys with sub-optimal assay concordance

For the surveys where agreement between the assays was suboptimal and/or there was no clear clustering of non-reactive and reactive samples, the effects of the model adjustments on the HIV prevalence estimates are case-specific.

In Burundi, the HIV prevalence estimate from the dichotomous model was similar to the unadjusted prevalence (1.9%). The continuous model, however, suggested that prevalence could be 0.1% point higher than the unadjusted estimate at 2.0% - but with a wide overlapping uncertainty interval. This small difference could result from the high number of samples that were positive on the first EIA and negative on the second EIA that were ultimately reactive according to the confirmatory line immunoassay.

In Niger, where all prevalence estimates overlap, the dichotomous model provides a slightly lower prevalence estimate (0.4%; 95% CrI: 0.3-0.5%) than the unadjusted estimates (0.5%; 95% CI: 0.4-0.7%). This resulted mainly from the combination of a low HIV prevalence and the relatively low estimated median posterior specificity for the first EIA in this country (99.3%; 95% CrI: 97.9-99.9%). The latter could have generated a high number of false positives. On the other hand, the results from the continuous model suggest that HIV prevalence could be 0.6% (95% CrI: 0.4-0.6%). In this case, prevalence is slightly higher because a high proportion of samples that were resolved with the line immunoassay are non-reactive or weakly reactive on the second EIA. As such, the continuous model gives a relatively high likelihood of being a true positive to samples that are close to the cutoff. For both dichotomous and continuous models, the low number of reactive samples makes it difficult to clearly estimate the distribution of true positives. In this context, our modeling assumptions will have more impact on HIV prevalence estimates.

In Cameroon and Gabon, both the dichotomous model and adjusted estimates, along with their confidence/credible intervals, were the same with an HIV prevalence of 4.6% (95% CrI: 4.2-4.9%) in Cameroon and 4.4% in Gabon (95% CrI: 4.0-4.8%). The estimates from the continuous model were slightly lower for both countries, however. In Cameroon and Gabon, the adjusted HIV prevalence estimates from the continuous model were 4.4% (95% CrI: 4.0-4.7%) and 4.2% (95% CrI: 3.8-4.6%), respectively. In both cases, the relatively high number of samples that were weakly reactive were given a higher likelihood of being true negatives than true positives.

In Haiti, the result from the dichotomous latent class model was the same as the unadjusted prevalence (2.3%; 95% CI: 2.1-2.5%), whereas the prevalence estimate from the continuous latent class model was slightly lower at 2.2% (95% CrI: 2.0-2.5%). Although this survey had censored OD values, the posterior variance of the true positive was rather small, which explained the slightly lower prevalence estimate. Our prevalence estimates for the Namibia DHS survey, which also had censored OD values, were relatively consistent between models: the unadjusted prevalence was 13.9% (95% CI: 13.2-14.7%), the prevalence from the dichotomous latent class model was 14.1% (95% CrI: 13.4-14.9%), while the prevalence from the continuous latent class model was 14.0% (95% CrI: 13.3-14.7%), with all uncertainty intervals overlapping.

In Sierra Leone, prevalence from the continuous latent class model suggested an estimate close to three times higher than the unadjusted and dichotomous latent class model. This is the consequence of the low posterior median of the distribution of true positives on the first EIA, which is very close to the cutoff. This low median resulted from the high proportion of samples initially negative on the first EIA that became positive on the second EIA. This result should be interpreted with caution because such a low median for the distribution of true positive on the first EIA could be considered unrealistic. More informative priors for the parameters governing the distribution of true positive could have resulted in lower prevalence.

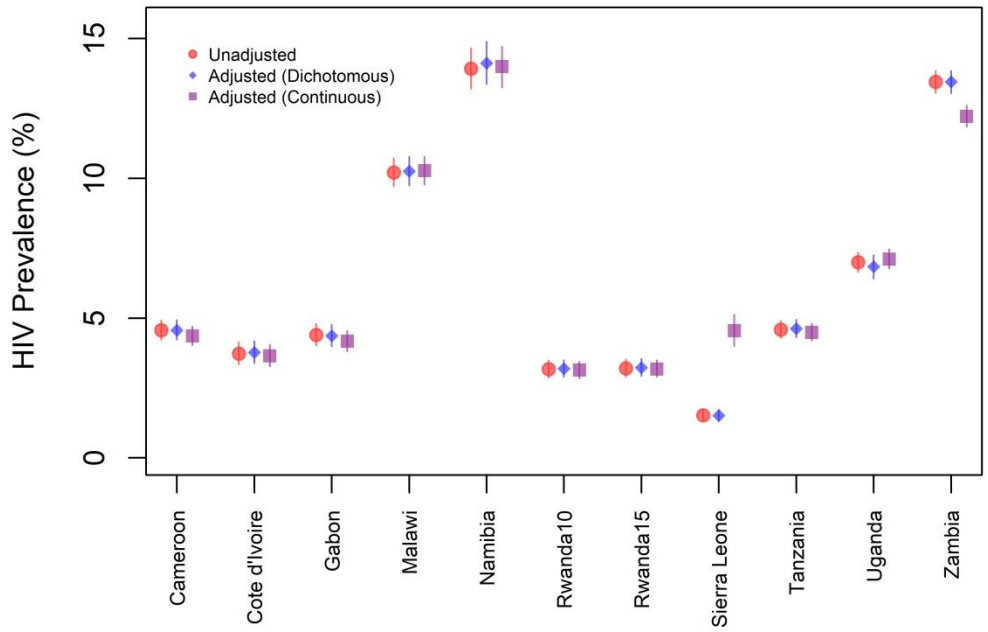

Figure 8A. Unadjusted and adjusted HIV prevalence estimates from Bayesian latent class models (high prevalence countries).

Unadjusted HIV prevalence estimates in countries where prevalence is above 3% along with the Bayesian latent class adjusted HIV prevalence estimates for the dichotomous and continuous models. Error bars correspond to the 95% confidence/credible intervals. (Note that estimates do not consider surveys weights and survey design.)

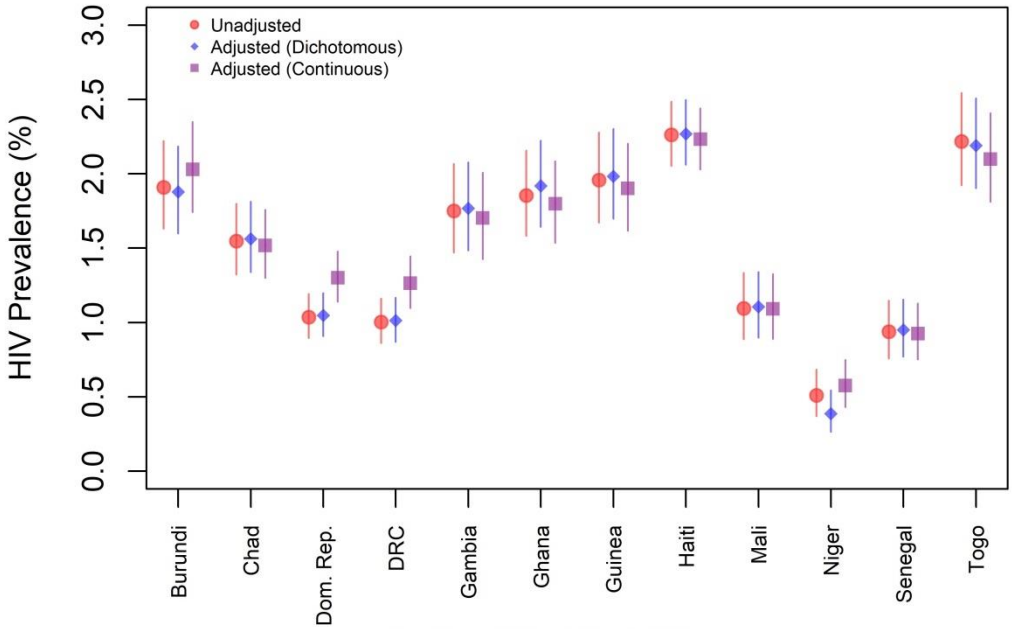

Figure 8B. Unadjusted and adjusted HIV prevalence estimates from Bayesian latent class models (low prevalence countries).

Unadjusted HIV prevalence estimates in countries where prevalence is below 3% along with the Bayesian latent class adjusted prevalence estimates for the dichotomous and continuous models. Error bars correspond to the 95% confidence/credible intervals. (Note that estimates do not consider surveys weights and survey design.)

Table 7. **Unadjusted versus adjusted HIV prevalence estimates (unweighted)**

The unadjusted and adjusted HIV prevalence estimates among men and women for the 23 surveys (age ≥15 years, upper age limit varies by survey). Adjusted prevalence estimates are provided for both the dichotomous and continuous latent class models.

	Unweighted HIV prevalence among men and women age ≥15 years (upper age limit varies by survey)		
Survey	**Unadjusted (95% CI)**	**Adjusted – Dichotomous test (95% CrI)**	**Adjusted- Continuous test (95% CrI)**
Sub Saharan Africa			
Burundi DHS 2010	1.9% (1.6-2.2%)	1.9% (1.6-2.2%)	2.0% (1.7-2.3%)
Cameroon DHS 2011	4.6% (4.2-4.9%)	4.6% (4.2-4.9%)	4.4% (4.0-4.7%)
Chad DHS 2014-15	1.5% (1.3-1.8%)	1.6% (1.3-1.8%)	1.5% (1.3-1.8%)
Côte d'Ivoire DHS 2011-12	3.7% (3.3-4.1%)	3.8% (3.4-4.2%)	3.7% (3.3-4.1%)
DRC DHS 2013-14*	1.0% (0.9-1.2%)	1.0% (0.9-1.2%)	1.0% (0.9-1.2%)
Gabon DHS 2012	4.4% (4.0-4.8%)	4.4% (4.0-4.8%)	4.2% (3.8-4.6%)
Gambia DHS 2013	1.8% (1.5-2.1%)	1.8% (1.5-2.1%)	1.7% (1.4-2.0%)
Ghana DHS 2014*	1.9% (1.6-2.2%)	1.9% (1.6-2.2%)	1.8% (1.6-2.2%)
Guinea DHS 2012*	2.0% (1.7-2.3%)	2.0% (1.7-2.3%)	1.9% (1.6-2.2%)
Malawi DHS 2010*	10.2% (9.7-10.7%)	10.3% (9.7-10.8%)	10.3% (9.8-10.8%)
Mali DHS 2012-13	1.1% (0.9-1.3%)	1.1% (0.9-1.3%)	1.1% (0.9-1.3%)
Namibia DHS 2013*	13.9% (13.2-14.7%)	14.1% (13.4-14.9%)	14.0% (13.3-14.7%)
Niger DHS 2012	0.5% (0.4-0.7%)	0.4% (0.3-0.5%)	0.6% (0.4-0.8%)
Rwanda DHS 2010*	3.2% (2.9-3.5%)	3.2% (2.9-3.5%)	3.1% (2.8-3.4%)
Rwanda DHS 2014-15*	3.2% (2.9-3.5%)	3.2% (2.9-3.5%)	3.2% (2.9-3.5%)
Senegal DHS 2010-11	0.9% (0.8-1.1%)	1.0% (0.8-1.2%)	0.9% (0.8-1.1%)
Sierra Leone DHS 2013	1.5% (1.3-1.7%)	1.5% (1.3-1.7%)	4.6% (4.0-5.1%)❖
Tanzania AIS/MIS 2011-12	4.6% (4.3-4.9%)	4.6% (4.3-4.9%)	4.5% (4.2-4.8%)
Togo DHS 2013-14	2.2% (1.9-2.5%)	2.2% (1.9-2.5%)	2.1% (1.8-2.4%)
Uganda AIS 2011	7.0% (6.7-7.3%)	6.8% (6.4-7.3%) 5.7% (4.7-6.5%)☆	7.1% (6.8-7.5%)
Zambia DHS 2013-2014*	13.4% (13.1-13.8%)	13.5% (13.1-13.9%)	12.2% (11.8-12.6%)
Latin America/Caribbean			
Dominican Republic DHS 2013*	1.0% (0.9-1.2%)	1.0% (0.9-1.2%)	1.0% (0.9-1.2%)
Haiti DHS 2012*	2.3% (2.1-2.5%)	2.3% (2.1-2.5%)	2.2% (2.0-2.5%)

*Surveys with censored optical density values

☆Using different priors, accounting for the potentially lower sensitivities and specificities of the three enzyme immunoassays used in the Uganda AIS 2013

❖Interpret this value with caution as the high number of samples tested as part of internal quality control that were found positive on the second enzyme immunoassay influences the distribution of true positive. Refer to the relevant results section for further details.

Surveys with no clear clustering of reactive EIA: Uganda and Zambia

The Uganda AIS 2011 survey only had 51.8% of initially reactive samples were reactive on the second EIA and did not include IQC. Furthermore, the third tie-breaker assay was also an EIA, which could have had a higher false positivity rate than if the third assay had been a Western Blot or line immunoassay.

Using the dichotomous model, we obtain an HIV prevalence estimate of 6.8% (95% CrI: 6.4-7.3%) in Uganda, slightly lower than the unadjusted estimate of 7.0% (95% CI: 6.7-7.3%) but with overlapping uncertainty intervals. Similarly, the continuous model estimated HIV prevalence at 7.1% (95% CrI: 6.8-7.5%). It is possible to validate these estimates by comparing them to prevalence estimates obtained with an algorithm of three rapid diagnostic tests (RDT) that were performed during field visits for that survey (Ministry of Health/Uganda and ICF International 2012). According to the RDT algorithm, the unweighted HIV prevalence in the Uganda AIS 2011 was 6.0% (95% CI: 5.7-6.3%). Using the Type 1 dichotomous latent class model presented in the methods section, we adjusted results from the RDT algorithm with estimates of sensitivity and specificity reported in the literature (not shown). The adjusted HIV prevalence estimates with the RDT algorithm were estimated at 5.7% (95% CrI: 5.4-6.1%). Comparing both EIA-based and RDT algorithms, we notice that the concordance of samples positive on the first EIA with the results from the RDT algorithm is rather poor (Figure 9). If the RDT algorithm is believed to be more accurate, this could suggest that the prevalence estimate that rely on the EIA-based algorithm could be overestimated in the Uganda AIS 2011.

It is possible to adjust, using the dichotomous model, the EIA-based algorithm to reproduce the 5.7% prevalence from the adjusted RDT algorithm. To do this, we need to provide priors with lower sensitivity and specificity for all three EIA. Specifically, if we assume that the 95% confidence intervals for the prior sensitivity is 90.0-99.0% (instead of 98.0-99.9%) and prior specificity is 85.0-95.0% (instead of 85.0-99.0%) for the three EIA, we obtain an adjusted prevalence estimate of 5.7% (95% CrI: 4.7-6.5%). This highlights the impact of our prior distributions for sensitivity and specificity on the adjusted prevalence estimates. This is especially true for the Uganda survey because of the absence of IQC. This means that the prior for the sensitivity of the first EIA will influence the prevalence estimates since there is no re-testing of potentially false negative samples. The lack of IQC also limited inferences from the continuous Bayesian latent class model and the true positive distribution of the second EIA was wide. The resulting posterior distribution of true positives on the third EIA was also very wide, with a median very close to the cutoff. This, in turn, increased prevalence estimates. Along with Sierra Leone, more informative priors on the distribution of true positives for Uganda could affect HIV prevalence estimates from this model.

To obtain a better sense of how HIV prevalence estimates from the EIA-based algorithm (unadjusted), RDT-based algorithm (unadjusted), and continuous Bayesian latent class adjustments compare in Uganda, we examine the sex-specific age trends (Figure 10). For both men and women, the RDT-based algorithm consistently yielded lower prevalence estimates across all age groups. The RDT-based algorithm yielded prevalence estimates that drop precipitously with age, with large drops between age 50-54 and 55-59 of 2.9 percentage points for men and by 4.1 percentage points for women, as compared to the unadjusted EIA-based algorithm. Differences between the EIA-based algorithm and the Bayesian adjustments were small for most age groups. The largest differences are observed for men in the 45-49 and 50-54 age groups, where the Bayesian adjustments were closer to the RDT-based algorithm. The relatively high HIV prevalence among young men with the EIA-based algorithm suggests that prevalence from the RDT-based algorithm is probably closer to the true HIV prevalence.

In contrast, the Zambia DHS 2013-14 survey included IQC and the concordance between assays was higher than in Uganda. In Zambia, however, HIV prevalence estimates could be affected by the very high proportion of weakly reactive samples on both EIA. Because the dichotomous model does not make any qualitative difference between weakly and strongly reactive specimens in its adjustment, the resulting

adjusted HIV prevalence estimates are not very different from the unadjusted ones. However, the adjusted prevalence estimate for the continuous model is lower at 12.2% (95% CrI: 11.8-12.6%), with a partly overlapping uncertainty interval. Here again, it is also possible to cross-validate these estimates with those of an independent confirmatory testing of some samples in Zambia. Recently, samples for the Zambia DHS 2013-14 survey that were classified as positives in the original testing algorithm - and which had one or two negatives RDT, or no RDT results at all - were independently re-tested with a line immunoassay (Inno-Lia HIV I/II). The samples that were positive on the original EIA-based algorithm and had positive results from the two RDT conducted in the household were considered positives without further testing with the line immunoassay. With this new algorithm, HIV prevalence decreased to 11.3% (95% CI: 10.5-11.9%) (unpublished data), which is lower than our adjusted estimates from the continuous model. There were, however, 632 samples with inconclusive test results on the confirmatory testing algorithm, meaning they were positive on the two EIA and negative or indeterminate on the line immunoassay. For the purpose of HIV prevalence estimation, these are considered negatives, although a proportion of them could be true positives.

A visual comparison of the confirmatory testing and the EIA-based results suggests that most weakly reactive samples on both EIA could be false positives (Figure 11). In fact, 97% (194/201) of samples that were weakly reactive on the first EIA – defined as having a S/CO value between 1 and 3 - and positive on the second EIA were found to be inconclusive according to the confirmatory testing algorithm including the line immunoassay. For samples that were positive on the first EIA and weakly reactive on the second EIA, 81% (311/385) were inconclusive in confirmatory testing. Notably, some samples that were highly reactive on both EIA were inconclusive in confirmatory testing. Overall, these observations suggest that unadjusted HIV prevalence estimates from the EIA-based algorithm could overestimate prevalence.

A further examination of the differences in estimates of HIV prevalence by age and sex in Zambia suggests that overestimation of prevalence in the EIA-based algorithm could be higher among men than for women (Figure 12). For women, the age trend in prevalence from the three types of prevalence estimates is relatively consistent. We note, however, that adjusted HIV prevalence estimates from the continuous Bayesian latent class model are generally slightly closer to the confirmatory testing results for younger females. For men, important differences are observed between the EIA-based algorithm and the confirmatory testing. The prevalence estimates from the adjusted continuous Bayesian latent class model are also much closer to those from the confirmatory testing than the EIA-based algorithm, especially for younger males (age 15-24). This could have resulted from a higher proportion (15%) of weakly reactive samples on both the first and second EIA of the EIA-based algorithm among young men. In contrast, only 5% of young women were weakly reactive on both the first and second EIA of the algorithm.

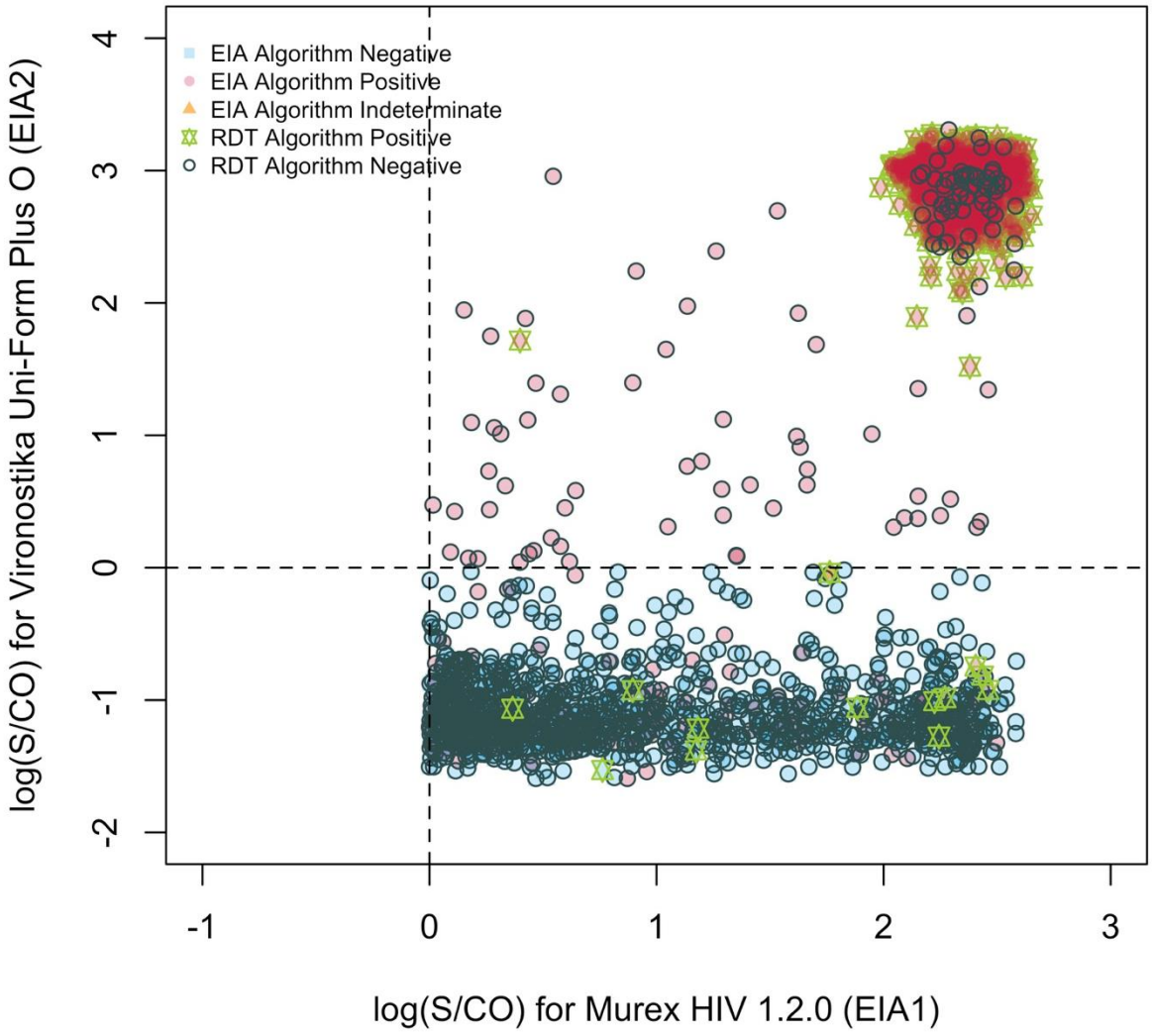

Figure 9. Comparison of the enzyme immunoassay-based and rapid test algorithms in the Uganda AIDS Indicator Survey (2011).

The HIV test results from the enzyme-immunoassay-based (EIA) and rapid diagnostic test-based (RDT) algorithm for sample in the 2011 Uganda AIS. Each possible final classification on both algorithms has its own symbol and color, where concordance between the two can be seen by the overlaying colors of the plot points.

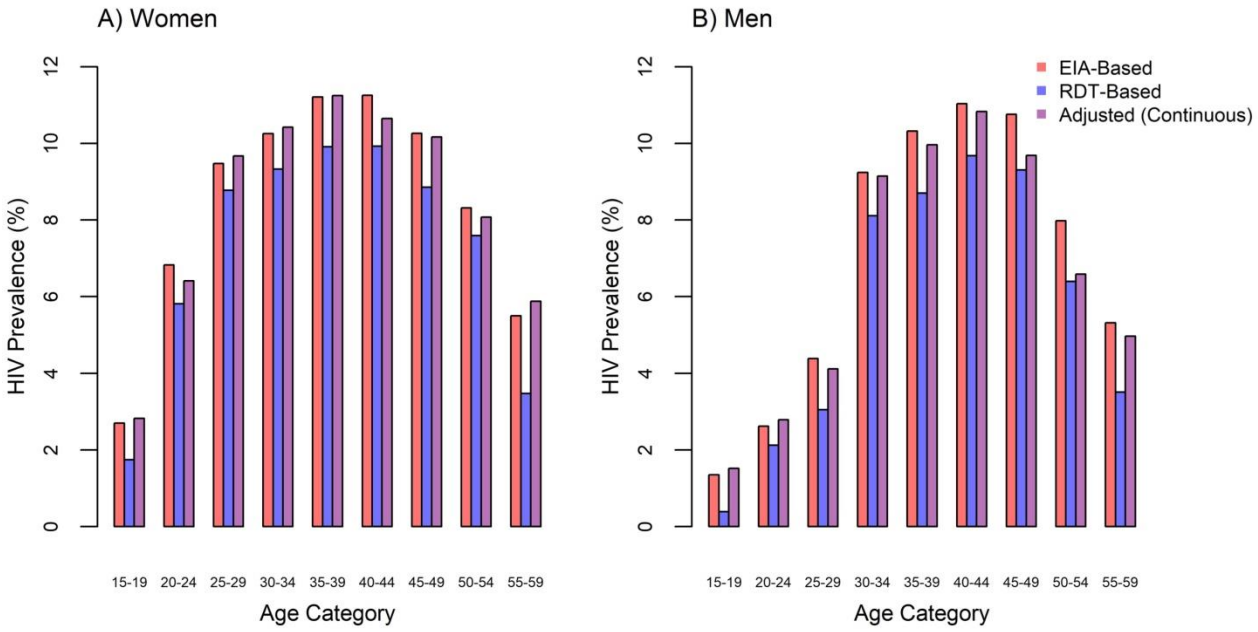

Figure 10. HIV prevalence from the Uganda AIDS Indicator Survey (2011) according to different testing strategies and adjustment procedures, stratified by age and sex.

Three different testing models are represented in the above graphs: EIA-based testing (unadjusted), rapid diagnostic test (unadjusted) and adjusted estimates from the continuous Bayesian latent class model. Data are stratified by sex and age.

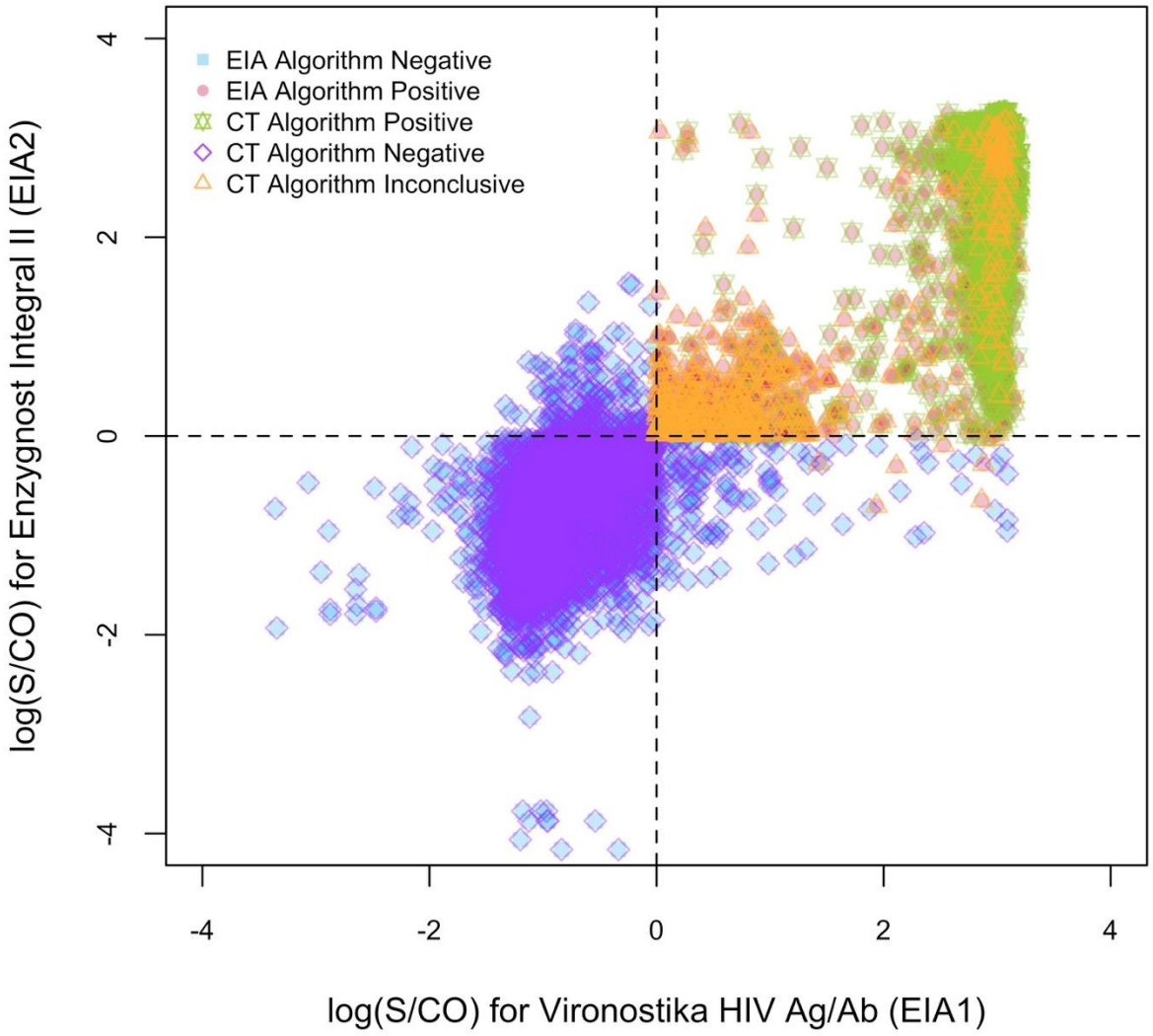

Figure 11. Comparison of the enzyme immunoassay-based algorithm with results from confirmatory re-testing in the Zambia Demographic and Health Survey (2013-14).

Graph displaying the HIV status classification for both the enzyme immunoassay-based (EIA) algorithm and confirmatory testing (CT) algorithm (using Inno-Lia HIV I/II). Specimens with positive results on the EIA and the RDT tests were considered confirmed positive and negative results on the EIA were considered confirmed negative with no further confirmatory testing with Inno-Lia. Samples were tested on the confirmatory testing algorithm if they were classified as HIV positive in the original laboratory testing algorithm (all samples in the upper right quadrant) and had one or two negative RDT, or no RDT results at all. For a sample to be considered positive on confirmatory testing, all 3 assays must be positive (green stars). If the first and second EIA were positive but the third assay was negative, the sample was classified as inconclusive (orange triangles). If the first and second EIA had discordant results, the final HIV status could be either inconclusive (if the confirmatory assay was positive, orange stars), negative (if the confirmatory assay was negative, purple diamonds) or indeterminate (if the confirmatory assay was indeterminate, but no samples were found to be indeterminate). Eight samples were excluded from the confirmatory testing algorithm due to insufficient blood from the original survey DBS collection (not shown on graph).

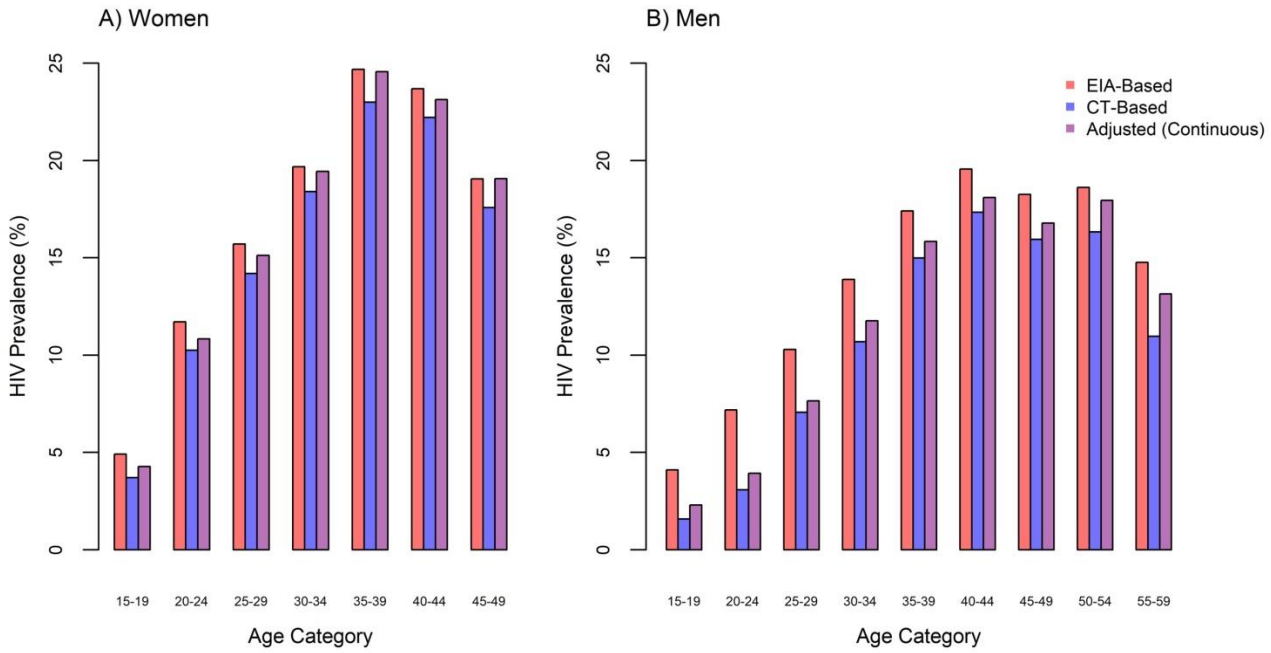

Figure 12. HIV prevalence from the Zambia Demographic and Health Survey (2013-14) according to different testing strategies and adjustment procedures, stratified by age and sex.

Three different testing models are represented in the above graphs: EIA-based testing (unadjusted), confirmatory testing (CT) (unadjusted), and the adjusted estimates from the continuous Bayesian latent class model. Data are stratified by sex and age.

4. Discussion

Using the Bayesian latent class models, we attempted to re-estimate results from the HIV seroprevalence surveys conducted by *The DHS Program* during 2010-2014. These adjustments account for the potential misclassification of HIV status that resulted from using a single cutoff value to qualitatively describe results as *positive* or *negative*. These analyses, along with visual inspection of the clustering of reactive and non-reactive samples, suggest that the misclassification of HIV status from the use of imperfect assays should not have a major influence on HIV prevalence estimates for most surveys. The observed differences between the adjusted and unadjusted results were often small and should be considered along with other potential biases such as random errors and/or selective non-response. Our analyses suggest, however, that EIA-based laboratory results from two surveys appear problematic: Uganda AIS 2011 and Zambia DHS 2013-14.

In the case of Uganda, we can triangulate our adjusted estimates with those from a RDT-based algorithm that was performed concomitantly with field data collection. Results from this RDT algorithm suggest that HIV prevalence could be 1% point lower (6.0% vs. 7.0%) than the EIA-based algorithm. Considering that errors in administering and interpreting results from RDT are probably less likely to occur than for the EIA-based algorithm (but not impossible), and that this latter algorithm resulted in low agreement between the first and second EIA, results from the EIA-based algorithm probably overestimate HIV prevalence. Because Uganda did not include IQC, our Bayesian latent class adjustments could be less precise than for other surveys and are more likely to be affected by our elicited prior distributions.

In Zambia, the high proportion of weakly reactive samples and the poor agreement between EIA raised concerns about potential overestimation of HIV prevalence. As expected, the dichotomous latent class model that does not consider the strength of the reaction did not produce a prevalence estimate that differed notably from the original unadjusted estimates. In this case, the continuous latent class model is more appropriate and yielded an HIV prevalence estimate of 12.2%, which was in between the unweighted estimates from the EIA-based algorithm (13.4%) and the confirmatory testing algorithm (11.3%). In this survey, the proportion of weakly reactive samples was highest among young males (age 15-24) and the EIA-based results could overestimate prevalence, especially within this group.

Limitations

Results from the dichotomous model are strongly influenced by the informative priors used for sensitivity and specificity. These priors reflect the high reported sensitivity and specificity of EIA, Western Blot, and line immunoassay when performed under good laboratory conditions. As exhibited by our comparisons between the EIA-based and rapid test algorithms in Uganda, the sensitivity and specificity of EIA can be much lower in practice. It is thus possible that the priors used for the dichotomous latent class models do not reflect the actual sensitivity and specificity of the EIA, as they were performed. If more realistic priors were used, the estimated prevalence might be quite different. Further work can study real-world conditions for the use of these assays to derive the best possible priors for the properties of the tests, which in turn should lead to more accurate estimates of the prevalence from these models.

The main assumption underlying the continuous latent class model is that the log(S/CO) values of the different EIA are normally distributed. In practice, however, this assumption could have been violated. For example, suboptimal laboratory practices (incorrect microplate washing, pipetting techniques, etc.) and performance of the assays (not validated for use on DBS) could have led to the misclassification of true negative samples as weakly reactive (i.e., false positive results). This would result in a skewed, non-normal distribution. We attempted to take this into account by imposing a minimum SD for the distribution of true negatives. This minimum SD would render the tails of the distribution “thicker”. It is nevertheless possible

that the true distribution of the log(S/CO) values within positive and negative populations is not normal. This is probably true for the Sierra Leone and Uganda surveys, where non-reactive and reactive samples do not seem to cluster and follow a multivariate normal distribution. In such cases, the results from this continuous Bayesian latent class model could be improved by using non-parametric models, but these have not yet been developed.

A further limitation needs to be acknowledged for surveys with censored OD values. For these surveys, the continuous latent class model imputes non-censored log(S/CO) values for each censored observation, and the resulting mean then depends on the allowed upper limit for the test values on the log scale. We attempted to mitigate this issue by imposing a reasonable maximum value for the mean and SD based on past information from these tests. The resulting posterior distribution of the true positives is affected by these priors which, although reasonable, influence HIV prevalence estimates.

5. Conclusions

Our examination of the impact of misclassification of HIV status on prevalence estimates has implications for population-based HIV surveillance. Importantly, it appears that for most surveys, the impact of misclassification due to testing errors is relatively minor compared to other potential sources of bias and uncertainty. We also found that the agreement between the first and second EIA is a relatively good indicator of the quality of laboratory testing, although it is not sufficient to rule out misclassification bias in HIV prevalence estimates. In addition, we suggest visual assessment of the clustering of non-reactive and reactive samples as a secondary measure of diagnostic accuracy.

The original EIA-based results in most surveys are considered reasonably accurate if we assume good assay properties. Our analyses suggest that this is most likely the case for surveys with good agreement between EIA and clear clustering of non-reactive and strongly reactive samples. In contrast, the EIA-based results from the Uganda AIS 2011 and the Zambia DHS 2013-14 surveys could be problematic. In Uganda, the lack of IQC and probable violation of the multivariate normality assumptions limits our confidence in results from the continuous Bayesian latent class model. For this survey, the true unobserved HIV prevalence is probably closer to results from the RDT-based rather than the EIA-based algorithm, as suggested by our examination of HIV trends by age and sex. In the case of Zambia, the EIA-based algorithm produced a high number of weakly reactive samples that were found later to be negatives on the confirmatory re-testing of samples. In this case, we recommend that future studies use the prevalence results from the confirmatory testing with the line immunoassay, although this algorithm could underestimate prevalence if the inconclusive test results from the line immunoassays that are strongly reactive on both EIA are indeed true positives.

In conclusion, our HIV prevalence estimates, adjusted for the imperfect sensitivity and specificity of the testing procedures using Bayesian latent class modeling, should be interpreted with caution because their underlying assumptions, including prior information crucial to some estimates, are difficult to verify in practice. Further knowledge about the test properties under real-world conditions will help to improve HIV prevalence estimates.

References

Abbott Laboratories. 2007. *Murex HIV 1.2.O* [Package insert]. United Kingdom: Abbott Laboratories.

Aboud, S., W. Urassa, E. Lyamuya, F. Mhalu, and G. Biberfeld. 2006. "Evaluation of HIV Antibody and Antigen/Antibody Combination ELISAS for Use in an Alternative Confirmatory HIV Testing Strategy in Dar Es Salaam, Tanzania." *J Virol Methods* 135(2):192-6.

Beelaert, G., G. Vercauteren, K. Fransen, M. Mangelschots, M. De Rooy, S. Garcia-Ribas, and G. van der Groen. 2002. "Comparative Evaluation of Eight Commercial Enzyme Linked Immunosorbent Assays and 14 Simple Assays for Detection of Antibodies to HIV." *J Virol Methods* 105(2):197-206.

Bio-Rad Laboratories. *2011. GS HIV Combo Ag/Ab EIA* [Package insert]. Redmond, Washington, USA: Bio-Rad Laboratories.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M.A. Brubaker, J. Guo, P. Li, and A. Riddell. 2016. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 20:1-37.

Centers for Disease Control and Prevention and Association of Public Health Laboratories (CDC and APHL). 2014. *Laboratory Testing for the Diagnosis of HIV Infection: Updated Recommendations.* CDC. Available at http://dx.doi.org/10.15620/cdc.23447.

Chacon, L., M. L. Mateos, and A. Holguin. 2017. "Relevance of Cutoff on a 4th Generation ELISA Performance in the False Positive Rate During HIV Diagnostic in a Low HIV Prevalence Setting." *J Clin Virol* 92:11-13.

Constantine, N. 2006. *HIV Antibody Assays.* UCSF Center for HIV Information. Available at http://hivinsite.ucsf.edu/InSite?page=kb-02-02-01#.

Fanmi, A. N., C. Ramière, J. C. Tardy, and P. Andre. 2013. "Real-Life Evaluation of a Human Immunodeficiency Virus Screening Algorithm Using a Single Combined P24 Antigen-Antibody Assay." *Eur J Clin Microbiol Infect Dis* 32(3):425-30.

Feng, C., H. Wang, N. Lu, T. Chen, H. He, Y. Lu, and X. M. Tu. 2014. "Log-Transformation and Its Implications for Data Analysis." *Shanghai Arch Psychiatry* 26(2):105-9.

Fishel, J.D., and D. Garrett. 2016. *Performance of Enzyme Immunoassays for* HIV *Serology in Surveys Conducted by the Demographic and Health Surveys Program. DHS Comparative Reports No. 39.* Rockville, Maryland: ICF International.

Francis, B. 2013. *Evaluation of the Bio-Rad Genscreen HIV Ag-Ab Ultra*. NRL. Available at http://nrl.gov.au/search/?value=genscreen&srcimg.x=0&srcimg.y=0.

Gray, R. H., F. Makumbi, D. Serwadda, T. Lutalo, F. Nalugoda, P. Opendi, G. Kigozi, S. J. Reynolds, N.K. Sewankambo, and M. J. Wawer. 2007. "Limitations of Rapid HIV-1 Tests During Screening

for Trials in Uganda: Diagnostic Test Accuracy Study." *BMJ : British Medical Journal* 335(7612):188-188.

Grüner, N., O. Stambouli, and R. S. Ross. 2015. "Dried Blood Spots - Preparing and Processing for Use in Immunoassays and in Molecular Techniques." *J Vis Exp* (97):52619.

Hakim, I., H. K. Patel, Y. T. Duong, J. Katoro, W. Gregory, I. Farouk, A. Hakim, and B. Parekh. 2016. *Evaluation of an EIA-Based Testing Algorithm in Antenatal Survey Using Dried Blood Spot Specimens from South Sudan.* Poster presented at the Conference on Retroviruses and Opportunistic Infections (CROI).

Hornik, K., F. Leisch, and A. Zeileis. 2003. *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. Paper in Proceedings of the Driving Simulation Conference (DSC), 2003.

Johnson, C., V. Fonner, A. Sands, S. Tsui, N. Ford, V. Wong, C. Overmeyer, and R. Baggaley. 2015. "Annex 14: A Report on the Misdiagnosis of HIV Status." In: *Consolidated Guidelines on HIV Testing Services: 5Cs: Consent, Confidentiality, Counselling, Correct Results and Connection 2015.* Geneva: WHO.

Joseph, L., T. W. Gyorkos, and L. Coupal. 1995. "Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard." *Am J Epidemiol* 141(3):263-72.

Lemee, V., M. Leoz, M. Etienne, F. De Oliveira, and J.C. Plantier. 2014. "Performance of the Liaison XL Murex HIV Ab/Ag Test on Clinical Samples Representing Current Epidemic HIV Variants." *J Clin Microbiol* 52(9):3277-9.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. "WinBUGS-a Bayesian Modelling Framework: Concepts, Structure, and Extensibility." *Statistics and Computing* 10(4):325-337.

MEASURE DHS. 2012. *MEASURE DHS Biomarker Field Manual*. ICF International. Calverton, Maryland, U.S.A.: ICF International.

Ministry of Health/Uganda and ICF International. 2012. *Uganda AIDS Indicator Survey (UAIS) 2011*. Calverton, Maryland, USA: Ministry of Health/Uganda and ICF International. Available at http://dhsprogram.com/pubs/pdf/AIS10/AIS10.pdf.

MP Diagnostics. 2003. *HIV Blot 2.2. Western Blot Assay* [Package Insert]. Geneva: MP Diagnostics.

Nimble - An R Package for Programming with BUGS Models Version 0.6-6. 2017.

Nkengasong, J. N., C. Maurice, S. Koblavi, M. Kalou, D. Yavo, M. Maran, C. Bile, K. N'Guessan, J. Kouadio, S. Bony, S. Z. Wiktor, and A. E. Greenberg. 1999. "Evaluation of HIV Serial and Parallel Serologic Testing Algorithms in Abidjan, Cote d'Ivoire." *AIDS* 13(1):109-17.

Parekh, B.S., and J. S. McDougal. 2005. "Application of Laboratory Methods for Estimation of HIV-1 Incidence." *Indian J Med Res* 121(4):510-8.

Peeling, R. W., H. Artsob, J. L. Pelegrino, P. Buchy, M. J. Cardosa, S. Devi, D. A. Enria, J. Farrar, D.J. Gubler, M. G. Guzman, S. B. Halstead, E. Hunsperger, S. Kliks, H. S. Margolis, C. M.

Nathanson, V. C. Nguyen, N. Rizzo, S. Vazquez, and S. Yoksan. 2010. "Evaluation of Diagnostic Tests: Dengue." *Nat Rev Microbiol* 8(12 Suppl):S30-8.

Piwowar-Manning, E., J. M. Fogel, P. Richardson, S. Wolf, W. Clarke, M. A. Marzinke, A. Fiamma, D. Donnell, M. Kulich, J. K. Mbwambo, L. Richter, G. Gray, M. Sweat, T.J. Coates, and S. H. Eshleman. 2015. "Performance of the Fourth-Generation Bio-Rad GS HIV Combo Ag/Ab Enzyme Immunoassay for Diagnosis of HIV Infection in Southern Africa." *J Clin Virol* 62:75-9.

Plummer, M. 2013. "rjags: Bayesian Graphical Models Using MCMC." *R package version* 3.

Qiu, X., S. Hodges, T. Lukaszewska, S. Hino, H. Arai, J. Yamaguchi, P. Swanson, G. Schochetman, and S. G. Devare. 2008. "Evaluation of a New, Fully Automated Immunoassay for Detection of HTLV-I and HTLV-II Antibodies." *J Med Virol* 80(3):484-93.

Sacks, R., A. Omodele-Lucien, N. Whitbread, D. Muir, and A. Smith. 2012. "Rapid HIV Testing Using Determine HIV 1/2 Antibody Tests: Is There a Difference between the Visual Appearance of True- and False-Positive Tests?" *Int J STD AIDS* 23(9):644-6.

Truant, A. L. 2016. *Manual of Commercial Methods in Clinical Microbiology, International Edition.* John Wiley & Sons.

UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance. 2005. *Guidelines for Measuring National HIV Prevalence in Population-Based Surveys*. Geneva: UNAIDS and WHO.

UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance. 2015. *Monitoring the Impact of the HIV Epidemic Using Population-based Surveys.* Geneva: UNAIDS and WHO.

Urassa, W., K. Godoy, J. Killewo, G. Kwesigabo, A. Mbakileki, F. Mhalu, and G. Biberfeld. 1999. "The Accuracy of an Alternative Confirmatory Strategy for Detection of Antibodies to HIV-1: Experience from a Regional Laboratory in Kagera, Tanzania." *J Clin Virol* 14(1):25-9.

Urio, L. J., M. A. Mohamed, J. Mghamba, A. Abade, and S. Aboud. 2015. "Evaluation of HIV Antigen/Antibody Combination ELISAs for Diagnosis of HIV Infection in Dar Es Salaam, Tanzania." *Pan Afr Med J* 20:196.

Wang, Z., N. Dendukuri, and L. Joseph. 2017. "Understanding the Effects of Conditional Dependence in Research Studies Involving Imperfect Diagnostic Tests." *Stat Med* 36(3):466-480.

Weber, B., L. Gurtler, R. Thorstensson, U. Michl, A. Muhlbacher, P. Burgisser, R. Villaescusa, A. Eiras, C. Gabriel, H. Stekel, S. Tanprasert, S. Oota, M. J. Silvestre, C. Marques, M. Ladeira, H. Rabenau, A. Berger, U. Schmitt, and W. Melchior. 2002. "Multicenter Evaluation of a New Automated Fourth-Generation Human Immunodeficiency Virus Screening Assay with a Sensitive Antigen Detection Module and High Specificity." *J Clin Microbiol* 40(6):1938-46.

World Health Organization (WHO). 1998. *Operational Characteristics of Commercially Available Assays to Determine Antibodies to HIV-1 and/or HIV-2 in Human Sera*. Geneva: WHO.

World Health Organization (WHO). 1999. *Operational Characteristics of Commercially Available Assays to Determine Antibodies to HIV-1 and/or HIV-2 in Human Sera.* Geneva: WHO.

World Health Organization (WHO). 2004. *HIV Assays: Operational Characteristics (Phase 1)*. Geneva: WHO.

World Health Organization (WHO). 2013. *WHO Prequalification of Diagnostics Programme. Product: Genscreen Ultra HIV Ag-Ab.* Geneva: WHO.

World Health Organization (WHO). 2015a. *Consolidated Guidelines on HIV Testing Services*. Geneva: WHO.

World Health Organization (WHO). 2015b. *HIV Assays. Laboratory Performance and Other Operational Characteristics. Rapid Diagnostic Tests (Combined Detection of HIV-1/2 Antibodies and Discriminatory Detection of HIV-1 and HIV-2 Antibodies)*. Geneva: WHO.

World Health Organization (WHO). 2015c. *WHO Prequalification of in Vitro Diagnostics Programme. Product: Murex HIV Ag/Ab Combination*. Geneva: WHO.

World Health Organization (WHO). 2016. *WHO Prequalification of in Vitro Diagnostics Programme - Public Report for MP Diagnostics HIV Blot 2.2*. Geneva: WHO.

Zeh, C., B. Oyaro, H. Vandenhoudt, P. Amornkul, A. Kasembeli, P. Bondo, D. Mwaengo, T. K. Thomas, C. Hart, K. F. Laserson, P. Ondoa, and J.N. Nkengasong. 2011. "Performance of Six Commercial Enzyme Immunoassays and Two Alternative HIV-Testing Algorithms for the Diagnosis of HIV-1 Infection in Kisumu, Western Kenya." *J Virol Methods* 176(1-2):24-31.

Zhang, R., Y. Sun, L. Wang, K. Zhang, J. Xie, and J. Li. 2013. "Blood Screening for Human Immunodeficiency Virus: A New Algorithm to Reduce the False-Positive Results." *Transfus Med* 23(4):260-4.

Appendix

Table S1. Target detection of the nine EIAs used in The DHS Program surveys (2010-2014)

The name, manufacturer, generation and format of the HIV screening enzyme immunoassays (EIA) used in the DHS reports. The solid phase antibodies and antigens coat the bottom of the microplate wells. The conjugate contains the substrate which allows for the detection of HIV antibodies or antigens in human sera or plasma.

			Solid Phase		Conjugate	
Assay Name (manufacturer)	**Generation**	**EIA format**	**Antibody detection**	**Antigen detection**	**Antibody detection**	**Antigen detection**
Enzygnost Integral II (Siemens)	4	Sandwich	R+S mixture: HIV-1 gp41, HIV-1 (subtype O) gp41, HIV-2 gp36	Anti p24 Mab	Biotin conjugated solid phase peptides	Biotin-conjugated anti-p24 Mab
Enzygnost Anti-HIV 1/2 Plus (Siemens)	3	Sandwich	HIV-1 gp41 (R) HIV-1 (subtype O) gp41 (R) HIV-2 gp36 (R)	-	(R) + (S) mixture of solid phase peptides	-
Genscreen Plus HIV Ag/Ab 5PL (Bio-Rad)	4	Sandwich	HIV-1 gp160 (R) HIV-1 (subtype O) gp41 (S) HIV-2 gp36 (S)	Anti p24 Mab	HIV-1 gp41 (R) HIV-1 (subtype O) gp41 (S) HIV-2 gp36 (R)	Biotin-conjugated polyclonal anti-p24
Genscreen ULTRA HIV Ag/Ab (Bio-Rad)	4	Sandwich	HIV-1 gp160 (R) HIV-1 (subtype O) gp41 (S) HIV-2 gp36 (S)	Anti p24 Mab	HIV-1 gp41 (R) HIV-1 (subtype O) gp41(S) HIV-2 gp36 (S)	Biotin-conjugated polyclonal anti-p24
HIV EIA (Ani Labsystems)	3	Sandwich	HIV-1 gp41 (S) HIV-1 gp120 (S) HIV-1 p24 (S) HIV-2 gp36 (S)	-	Same as solid phase	-
Murex HIV 1.2.0 (Abbott Labs)	3	Sandwich	HIV-1 gp41 (R) HIV-1 p24 (R) HIV-1 (subtype O) gp41 (S) HIV-2 gp36 (R)	-	Same as solid phase	-
Murex HIV Ag/Ab (DiaSorin)	4	Sandwich	HIV-1 gp41 (R) HIV-1 *polymerase* (R) HIV-1 (subtype O) gp41 (R) HIV-2 gp36 (R)	Anti p24 Mab	Same as solid phase	HRP conjugated anti-p24 Mab
Vironostika HIV Ag/Ab (Biomérieux)	4	Sandwich	HIV-1 gp160 (R) HIV-1 p24 (R) HIV-1 (subtype O) gp41 (S) HIV-2 gp36 (S)	Anti-p24 Mab	Same as solid phase	HRP conjugated anti-p24 Mab
Vironostika HIV Uni-Form 2 Plus O (Biomérieux)	3	Sandwich	HIV-1 gp160 (R) HIV-1 p24 (R) HIV-1 (subtype O) gp41 (S) HIV-2 gp36 (S)	-	Same as solid phase	-

Mab: monoclonal antibody, (R): recombinant peptide, (S): synthetic peptide

Table S2. Target detection of supplementary assays

The name, manufacturer, and type of each supplementary assay used in the DHS reports, with their respective antigen detection profile

		Antibody detection in human sera or plasma																				
		HIV-1 antigens											HIV-2 antigens									
Assay (manufacturer)	**Assay type**	**gp 41**	**gp 120**	**gp 160**	**p17**	**p18**	**p24**	**p31**	**p39**	**p51**	**p55**	**p66**	**gp 36**	**gp 105**	**gp 125**	**gp 140**	**p16**	**p26**	**p34**	**p56**	**p68**
Inno-Lia HIV I/II Line Score (Innogenetics)	Line immuno-assay	X	X		X		X	X				X	X	X							
New Lav Blot (Bio-Rad)*	Western Blot	X	X	X		X	X	X		X	X	X	X	X	X	X	X	X	X	X	X
HIV Blot 2.2 (DiaSorin) Western Blot 2.2 (Abbott Labs)	Western Blot	X	X	X	X		X	X	X	X	X	X	X								

*The New Lav Blot assay uses two separate strips to detect both HIV types.

Text S1: Justification of prior distributions for the Bayesian latent class model's parameters

A) Priors for the continuous assays (EIA)

Enzygnost Integral II (Siemens)

The mean S/CO ratio of confirmed positive samples collected in Dar es Salaam (Tanzania) using Enzygnost Integral II was independently estimated at 13.7 (95% CI: 13.1-14.5) (Urio et al. 2015). Because this study was conducted in Tanzania and other countries could have different average values, the standard deviation of this parameter was chosen so that this mean has a 99% probability of being higher than a S/CO value of 5 (threshold for highly reactive specimens). This translates in the following prior for the mean of the distribution of true positive (v_{t1}):

$v_{t1} \sim Normal\ (log(13.7),\ 0.4333))$

For the standard deviation of the distribution of true positives (σ_{t1}), the following non-informative prior was chosen:

$\sigma_{t1} > 0$

For the distribution of true negatives, the mean S/CO value of confirmed HIV negatives was not available from the literature. To provide a prior for this assay, we examined the median S/CO values of the surveys with the two highest concordances between the first and second EIA: 0.25 in Côte d'Ivoire and 0.38 in Chad (for a log-average of 0.31). We estimated the SD of this parameter so that there is a 95% probability that the mean lies within the 0.25-0.38 range (the range of values given by the two surveys with highest concordance). Accordingly, we will provide the following prior for the mean of the distribution of true negatives (v_{t0}):

$v_{t0} \sim Normal(log(0.31),\ 0.1097)$

For the standard deviation of this distribution (σ_{t0}), we chose a minimum SD so that the probability that a true negative has a value greater than the cutoff is 5% (conditional on a mean S/CO of 0.31 for true negatives). This reflects the fact that most weakly reactive samples are probably due to lab errors and/or higher false positivity resulting from the use of DBS.

$\sigma_{t0} \geq 0.7120$ *(note this is the SD for the log-transformed S/CO)*

Enzygnost Anti-HIV 1/2 Plus (Siemens)

Two independent studies were conducted in Tanzania and reported average S/CO values of 7.5 (95 CI%: 7.3-7.7) and 8.0 (95% CI: 7.8-8.2) for this test (Aboud et al. 2006; Urassa et al. 1999). Averaging this value (log-scale), we obtain 7.7. Because these two studies were conducted in Tanzania and other countries could have different average values, the standard deviation of this parameter was chosen so that this mean has a 99% probability of being higher than a S/CO value of 5 (threshold for highly reactive specimens). This translates in the following prior for the mean of the distribution of true positive (v_{t1}):

$v_{t1} \sim Normal(log(7.7),\ 0.1856))$

For the SD of the distribution of true positives (σ_{t1}), we chose a non-informative prior, such that:

$\sigma_{t1} > 0$

For the distribution of true negatives, the mean S/CO value of confirmed HIV negatives was not available from the literature. To provide a prior for this assay, we examined the distributions of S/CO values of the survey with the highest concordance between the first and second EIA. This would correspond to Malawi which reached 99.0% concordance after repeated testing (including IQC). In this survey, the median S/CO values of negative sample were estimated at 0.23. It is difficult to estimate a prior for this estimate's precision and we gave it a weakly informative prior. Specifically, the precision was chosen so that there is a 95% probability that the mean lies in the 0.10-0.53 interval.

$v_{t0} \sim Normal(log(0.23), 0.4250))$

For the standard deviation of this distribution (σ_{t0}), we chose a minimum SD such that the probability that a true negative has a value greater than the cutoff is 5% (conditional on a mean S/CO of 0.23 for true negatives). This reflects the fact that most weakly reactive samples are probably due to lab errors and/or higher false positivity resulting from the use of DBS.

$\sigma_{t0} \geq 0.8935$ *(note this is the SD for the log-transformed S/CO)*

Genscreen Plus HIV Ag/Ab 5PL (Bio-Rad)

There is no information available in the literature on the distribution of S/CO values for true positives and negatives for this assay. Only one survey used this test (Gabon) and agreement between the two EIA in this survey was suboptimal. The prior for the distribution of true positives was assumed to have a mean S/CO of 15 and the standard deviation chosen so that the parameter has a 99% probability of being higher than 5 (threshold for highly reactive specimens). This translates in the following prior for the mean of the distribution of true positive (v_{t1}):

$v_{t1} \sim Normal(log(15), 0.4722))$

For the SD of the distribution of true positives (σ_{t1}), we chose a non-informative prior, such that:

$\sigma_{t1} > 0$

For the distribution of true negatives, the mean S/CO value of confirmed HIV negatives was not available from the literature. The prior for the mean (v_{t0}) of this distribution was thus assumed to be 0.30, with a 95% probability of lying in the 0.15-0.60 interval.

$v_{t0} \sim Normal(log(0.30), 0.3537)$

For the standard deviation of this distribution (σ_{t0}), we chose a minimum SD such that the probability that a true negative has a value greater than the cutoff is 5% (conditional on a mean S/CO of 0.30 for true negatives).

$\sigma_{t0} \geq 0.7320$ *(note this is the SD for the log-transformed S/CO)*

Genscreen ULTRA HIV Ag/Ab (Bio-Rad)

Only one survey (Cameroon) used this EIA and there are no data in the literature on the distribution of S/CO for confirmed positives and negatives. Hence, the same general priors used for Greenscreen Plus HIV Ag/Ab 5 PL are used.

HIV EIA (Ani Labsystems)

Only one survey (Uganda) used this EIA and there are no data in the literature on the distribution of S/CO for confirmed positives and negatives. The same general priors used for Greenscreen Plus HIV Ag/Ab 5 PL and GreenScreen ULTRA HIV Ag/Ab are used.

Murex HIV 1.2.0 (Abbott Labs)

Only one survey (Uganda) used this EIA and there are no data in the literature on the distribution of S/CO for confirmed positives and negatives. The same general priors used for Greenscreen Plus HIV Ag/Ab 5 PL, GreenScreen ULTRA HIV Ag/Ab, and HIV EIA are used.

Murex HIV Ag/Ab combination (Diasorin)

The mean S/CO ratio of confirmed positives samples was estimated at 12.2 and 13.3 in two independent studies (Aboud et al. 2006; Urio et al. 2015). The mean for the distribution of true positives was chosen as the average (log scale) of these two studies. The standard deviation of this parameter was chosen so that this mean has a 99% probability of being higher than a S/CO value of 5 (threshold for highly reactive specimens). This translates in the following prior for the mean of the distribution of true positive (v_{t1}):

$v_{t1} \sim Normal(log(12.7), 0.4007))$

For the SD of the distribution of true positives (σ_{t1}), we chose a non-informative prior, such that:

$\sigma_{t1} > 0$

For the distribution of true negatives, the mean S/CO value of confirmed HIV negatives was not available from the literature and the only survey that used this assay (Cameroon). In Cameroon, this first EIA did not have good concordance with the other assay. (Rwanda 2010 and Rwanda 2014-15 also used this test but as the second EIA.) As for previous assays, the prior for the mean of this distribution (v_{t0}) was thus assumed to be 0.3 with a 95% probability of lying in the 0.15-0.60 interval.

$v_{t0} \sim Normal(log(0.30), 0.3537)$

For the standard deviation of this distribution (σ_{t0}), we chose a minimum SD such that the probability that a true negative has a value greater than the cutoff is 5% (conditional on a mean S/CO of 0.30 for true negatives).

$\sigma_{t0} \geq 0.7320$ *(note this is the SD for the log-transformed S/CO)*

Vironostika HIV Ag/Ab (Biomérieux)

The mean S/CO ratio of confirmed positives samples was estimated at 15.5 and 15.8 in two independent studies (Aboud et al. 2006; Urio et al. 2015). Surveys with the highest concordance on the first and second EIA that used Vironostika HIV Ag/Ab as the first assay – and whose OD distributions were not censored – generally had higher median S/CO: 24.9 (Chad), 25.7 (Togo), and 29.8 (Tanzania). The prior for the mean of this value was chosen to correspond to the midpoint between the minimum (15.5) and maximum (29.8) of these values (the midpoint on the log scale of 21.5). The standard deviation of this parameter was chosen

so that this mean has a 99% probability of being higher than a S/CO value of 5 (threshold for highly reactive specimens). This translates in the following prior for the mean of the distribution of true positive (v_{t1}):

$v_{t1} \sim Normal(log(21.5), 0.6270))$

For the SD of the distribution of true positives (σ_{t1}), we chose a non-informative prior, such that:

$\sigma_{t1} > 0$

For the distribution of true negatives, we use the mean S/CO of the negatives samples of the Chad survey, which had the best concordance (99.8% concordant) between the two EIA. This mean was 0.33. Other surveys which use the same firs assay and that had good concordance (>98.5%) had similar means (Togo 0.33, Dominican Republic 0.36, Tanzania 0.35, Rwanda 2010 0.40, Guinea 0.53, Ghana 0.31). It is difficult to estimate a prior for this estimate's precision and we gave it a weakly informative prior with 95% probability of lying in the 0.15-0.54 interval.

$v_{t0} \sim Normal(log(0.33), 0.2555)$

For the standard deviation of this distribution (σ_{t0}), we chose a minimum SD such that the probability that a true negative has a value greater than the cutoff is 5% (conditional on a mean S/CO of 0.33 for true negatives).

$\sigma_{t0} \geq 0.6740$ *(note this is the SD for the log-transformed S/CO)*

Vironostika HIV Uni-Form II Plus O (Biomérieux)

The mean S/CO ratio of confirmed positives samples was estimated in an independent study at 14.8 in Tanzania (Aboud et al. 2006). Among surveys with the best concordance on first and second EIA, and that used this test on the first assay, the median S/CO was highest at 25.5 in Mali and 21.7 in Senegal. The prior for the mean of this parameter was chosen to correspond to the midpoint between the minimum and maximum of these values (the midpoint on the log scale of 19.4). The standard deviation of this parameter was chosen so that this mean has a 99% probability of being higher than a S/CO value of 5 (threshold for highly reactive specimens). This translates in the following prior for the mean of the distribution of true positive (v_{t1}):

$v_{t1} \sim Normal(log(19.4), 0.5828))$

For the SD of the distribution of true positives (σ_{t1}), we chose a non-informative prior, such that:

$\sigma_{t1} > 0$

For the distribution of true negatives, we use the mean S/CO of the negatives samples of the two surveys (Mali and Senegal) which had the best concordance (>97%) between the two EIA. The mean S/CO of the negative samples was estimated at 0.34 and 0.32 in Mali and Senegal, respectively. Because this median is very close to the mean of negative samples tested with the 4th generation version of this test, the same parameters were used. The prior for the mean of the true negative (v_{t0}) is:

$v_{t0} \sim Normal(log(0.33), 0.2555)$

For the standard deviation of this distribution (σ_{t0}), we chose a minimum SD such that the probability of that a true negative has a S/CO value greater than the cutoff is 5% (conditional on a mean S/CO of 0.33 for

true negatives). This reflects the fact that most weakly reactive samples are probably due to lab errors and/or higher false positivity resulting from the use of DBS.

$\sigma_{t0} \geq 0.6740$ *(note this is the SD for the log-transformed S/CO)*

B) Priors for dichotomous assays (Western Blot or line immunoassay)

Inno-Lia HIV I/II Score line immunoassay (Innogenetics)

The sensitivity and specificity of this test are estimated to be 100% (95% CI: 98.6%-100%) and 100% (95% CI: 98.0-100%), respectively (WHO 2015b). For the posterior distributions of this test's sensitivity and specificity to remain within these ranges, the lower limit of the confidence interval defined the prior's lower bound. Specifically, we have:

$Se_w \sim Beta(1,1)*I(0.986, 1)$

$Sp_w \sim Beta(1,1)*I(0.980, 1)$

New Lav Blot (Bio-Rad)

The sensitivity and specificity of this test are estimated to be 100% (95% CI: 98.1%-100%) and 100% (95% CI: 96.8-100%), respectively (WHO 2015b). For the posterior distributions of this test's sensitivity and specificity to remain within these ranges, the lower limit of the confidence interval defined the prior's lower bound.

$Se_w \sim Beta(1,1)*I(0.981, 1)$

$Sp_w \sim Beta(1,1)*I(0.968, 1)$

HIV 2.2 Blot (DiaSorin) / Western Blot 2.2 (Abbott Labs)

The sensitivity and specificity of this test are estimated by the test manufacturer, excluding indeterminate samples, to be 100% (95% CI: 98.2%-100%) and 100% (95% CI: 98.0-100%) for HIV-1, respectively (WHO 2016). For the posterior distributions of this test's sensitivity and specificity to remain within these ranges, the lower limit of the confidence interval defined the prior's lower bound.

$Se_w \sim Beta(1,1)*I(0.982, 1)$

$Sp_w \sim Beta(1,1)*I(0.980, 1)$

Text S2: Details on Bayesian Latent Class Continuous Model Specification

Table S3 below summarizes the different patterns of tests used for classifying the different algorithm for the Bayesian latent class analysis based on the distribution of the continuous assay data. All countries, except Uganda, use two continuous diagnostic tests. "Type 1" surveys administer one test to all subjects; those who tested positive on the first test are then administered a second test (along with those subjects selected for IQC). If the two test results disagree, a final dichotomous test is used.

For "Type 2" and "Type 3" surveys, the first and second EIA are repeated in parallel if their results do not match. For "Type 2" surveys, a final dichotomous test is administered to subjects for whom the repeated tests still disagree (for "Type 3" survey, there is no third assay).

Uganda is the only "Type 4" survey. A first assay is performed on all samples, a second EIA is administered to subjects who test positive to the first assay, and a third EIA is conducted on specimens with discrepant results on the first two EIA. In surveys with a Type 4 testing strategy, there is no confirmatory dichotomous assay such a Western Blot or line immunoassay.

Table S3 Data Structure

Survey type	Description	Subjects sets	Parameters list (data sets)	
Type 1	One or two continuous assay (only once each) are used	S_1, S_2	μ_2, σ_2, ρ μ_1, σ_1	S_2 S_2 & S_1
Type 2 and 3	One or two continuous assay & parallel testing are used	S_1, S_2, S_4	ρ_{11}, ρ_{22} $\mu_2, \sigma_2, \rho_{12}$ μ_1, σ_1	S_4 S_4 & S_2 ($\rho = \rho_{12}$) S_4, S_2 & S_1
Type 4	Up to three continuous assays (only once each) are used	S_1, S_2, S_3	$\mu_3, \sigma_3, \rho_{13}, \rho_{23}$ $\mu_2, \sigma_2, \rho_{12}$ μ_1, σ_1	S_3 S_3 & S_2 ($\rho = \rho_{12}$) S_3, S_2 & S_1

MCMC algorithm

We wrote our custom Monte Carlo Markov Chain (MCMC) algorithm to sample the posterior distributions of interest. This code was written in R where we generated random values for each parameter based on their respective full conditional posterior distributions at each iteration.

Developing the likelihood terms L_{Xi}

The likelihood terms $L_{i1|s}$ - $L_{i4|s}$ are involved in the calculation of L_{Xi} which are of the form:

$$L_{ik|s} \propto \frac{1}{|\Sigma|}\exp -\frac{1}{2}\{(X_i - \mu)^\top \Sigma^{-1}(X_i - \mu)\}$$
$$= \frac{1}{|\Sigma|}\exp -\frac{1}{2|\Sigma|}\{(X_i - \mu)^\top \Lambda(X_i - \mu)\}$$
$$= \frac{1}{d}\exp -\frac{q}{2d}$$

Where Λ is the cofactor matrix of Σ, d is its determinant, $d = |\Sigma|$and $q = (X_i - \mu)^\top \Lambda (X_i - \mu)$. Note that the quadratic term q above is a polynomial of degree 2 in terms of μ_1, μ_2 and/or μ_3.

It is not important to acknowledge the following at this point, but q is also a polynomial of degree 2 in terms of the different correlation terms ($\rho's$), while it is a polynomial of degree 2 or higher in terms of SD's ($\sigma's$).

Full conditional posterior distributions for the means

The full conditional posterior distribution for the mean of assay t scores in either the positive (s = 1) or negative (s = 0) assay distributions (μ_{ts}) is relatively easy to drive since μ_{ts} appears only in the quadratic parts q of the various likelihood terms.

We illustrate its calculation through the elicitation of μ_2's full conditional posterior distribution. We drop the s index indicating to which distribution the corresponding mean corresponds: both cases (s=0, 1) are done exactly along the same lines.

Depending on the survey type, μ_2 is present in L_2 (Types 1 and 2), L_2 and L_3 (Type 4) or L_2, L_3 and L_4 (Type 3). Again, we illustrate only the calculation of the term in μ_2 present in L_4 for the sake of conciseness, as the contribution of terms L_2 and L_3 to μ_2's full conditional posterior distributions are derived the same way.

The contribution of subject $i \in S_4$ to μ_{2s} full conditional posterior distribution (provided the true latent status of subject i is equal to s, that is, $T_i = s$) is proportional to:

$$L_{4i} \propto \exp\left\{-\frac{1}{2}(\mathrm{X_i} - \mu_\mathrm{s})^\top \Sigma_{4s}^{-1}(\mathrm{X_i} - \mu_\mathrm{s})\right\}$$

Where μ_2 is absent of the covariance matrix Σ_{4s} and the mean is:

$$\mu_s = \begin{pmatrix}\mu_{1s}\\ \mu_{1s}\\ \mu_{2s}\\ \mu_{2s}\end{pmatrix} = \underbrace{\mu_{1s}\begin{pmatrix}1\\1\\0\\0\end{pmatrix}}_{K_1} + \underbrace{\mu_{2s}\begin{pmatrix}0\\0\\1\\1\end{pmatrix}}_{K_2}$$

Hence the contribution of subjects in S_4 to μ_{2s} full conditional posterior distribution is proportional to:

$$\prod_{i\in S_4:\, T_i=s} L_{4i} \propto \exp\left\{-\frac{1}{2}\sum_{i\in S_4:\, T_i=s}\left(\underbrace{X_i - \mu_{1s}K_1}_{\tilde{X}_i} - \mu_{2s}K_2\right)^\top \Sigma_{4s}^{-1}\left(X_i - \mu_{1s}K_1 - \mu_{2s}K_2\right)\right\}$$

$$= \exp\left\{-\frac{1}{2}\sum_{i\in S_4:\, T_i=s}\left(\tilde{X}_i^\top - \mu_{2s}K_2{}^\top\right)\Sigma_{4s}^{-1}\left(\tilde{X}_i - \mu_{2s}K_2\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\mu_{2s}^2 \underbrace{n_{4s}K_2^\top \Sigma_{4s}^{-1}K_2}_{A_4} - 2\mu_{2s}\underbrace{\sum_{i\in S_4:\, T_i=s} K_2^\top \Sigma_{4s}^{-1}\tilde{X}_i}_{B_4}\right)\right\}$$

Where n_{4s} is the number of subjects in S_4 with status s, $s = 0, 1$. Similarly, subjects in S_2 contribute to the full conditional posterior distribution of μ_{2s} and we note (A_2, B_2) the terms corresponding to (A_4, B_4) in the above expression. Since the prior distribution for to μ_{2s} is Normal (to $\mu_{02s}, \sigma^2_{02s}$), we have:

$$f(\mu_{2s}) \propto \exp\left\{-\frac{1}{2\sigma^2_{02s}}(\mu_{2s} - \mu_{02s})^2\right\}$$
$$\propto \exp\left\{-\frac{1}{2}\left(\mu^2_{2s}\frac{1}{\sigma^2_{02s}} - 2\mu_{2s}\frac{\mu_{02s}}{\sigma^2_{02s}}\right)\right\}$$

It follows that the full posterior conditional distribution is of the form:

$$f(\mu_{2s}|\, data, other\ parms) \propto \exp\left\{-\frac{1}{2}(\mu^2_{2s}A - 2\mu_s B)\right\}$$
$$where\ A = \sum_{j=0,2,4} A_j$$
$$and\ B = \sum_{j=0,2,4} B_j$$

That is, μ_{2s} full conditional posterior distribution is also normal. If a constraint was imposed on the maximum value for the mean, the posterior distribution is the same as the one below but truncated to the limits imposed.

$$\mu_{2s}|\, data, other\ parms \sim N\left(mean = \frac{B}{A}, variance = \frac{1}{A}\right)$$

Full conditional posterior distributions for correlations and SDs

The elicitation of SD's or correlations' full conditional posterior distributions is by far the most complex in this project. However, their calculations are manageable after we recognize their form.

Consider for example the full conditional posterior distribution of σ_2 in a survey of Type 3, that is, a survey with two EIA and parallel retesting of discrepant results. As we can see from Table S3, the parameter σ_2 is present in the multivariate normal distribution for subjects in either S_2 or S_4.

The multivariate normal distribution for subjects in S_2 (or S_4) is presented and worked out - in terms of $\theta \in \{\rho_{11}, \rho_{12}, \rho_{22}, \sigma_1, \sigma_2\}$'s full conditional posterior distribution. In Type 3 surveys, the log of the full conditional posterior distribution for σ_2 is given by:

$$l(\sigma_2|data, other\ parameters)$$
$$= l_2(\sigma_2|data, other\ parameters) + l_4(\sigma_2|data, other\ parameters)$$

where l_2 and l_4 are obtained from the equation above, applied to subjects from S_2 and S_4 respectively. Its first two derivatives are the sum of the derivatives of l_2 and l_4 (respectively, the first- and second-order), also obtained from the equations above.

Full conditional posterior distribution for the latent *True Status* T_i

At each iteration, the *True Status* of subject i, T_i, is updated/resampled. Note that T_i appears in the likelihood terms L_{Ti}, L_{Xi}, and L_{Wi} when subject i was administered a dichotomous test (Western Blot or line immunoassay). Hence, the full conditional posterior distribution for T_i; i = 1, 2, ..., N can be written as a product of terms exponentiated with either T_i or 1-T_i, that is

$$f(T_i|\, X_i, \theta) \propto \left(\prod_j a_{ij}\right)^{T_i} \left(\prod_j b_{ij}\right)^{1-T_i}$$
$$= a_i^{T_i} b_i^{1-T_i}$$
$$\Rightarrow T_i|\, X_i, \theta \sim \text{Bernoulli}\left(\frac{a_i}{a_i + b_i}\right)$$

Full conditional posterior distribution for the prevalence

The full conditional posterior distribution for prevalence is the easiest to derive, because prevalence is only present in the likelihood component L_{Ti}. Its posterior distribution is thus given by:

$$f(\pi \mid X, \theta_{-\pi}) \propto \prod_i \pi^{T_i} (1-\pi)^{1-T_i}\, \pi^{\alpha_\pi - 1}(1-\pi)^{\beta_\pi - 1}$$
$$= \pi^{\sum T_i + \alpha_\pi - 1}(1-\pi)^{N - \sum T_i + \beta_\pi - 1}$$
$$\sim Beta\left(\sum T_i + \alpha_\pi, N - \sum T_i + \beta_\pi\right)$$

Full conditional posterior distributions for the sensitivities and specificities of the dichotomous tests

In surveys where a dichotomous test was used, we update its sensitivity and specificity values by sampling values from their respective full conditional posterior distributions at each iteration. Since these tests contribute to the likelihood term and their parameters' prior distributions are Beta distributions, their respective full conditional posterior distributions are also Beta distributions. In cases where a truncated prior is used, the full conditional posterior density will be a truncated beta density, that is, a beta density over a range narrower than [0, 1]. The full conditional posterior distributions are:

$$f(s_W \mid data, other\ parameters) \propto \prod_i \left[s_W^{w_i}(1-s_W)^{1-w_i}\right]^{T_i} s_W^{\alpha_s - 1}(1-s_W)^{\beta_s - 1}$$
$$= s_W^{\sum_i T_i w_i + \alpha_s - 1}(1-s_W)^{\sum_i T_i (1-w_i) + \beta_s - 1}$$
$$and\ f(c_W \mid data, other\ parameters) \propto \prod_i \left[(1-c_W)^{w_i} c_W^{1-w_i}\right]^{(1-T_i)} c_W^{\alpha_c - 1}(1-c_W)^{\beta_c - 1}$$
$$= c_W^{\sum_i (1-T_i)(1-w_i) + \alpha_c - 1}(1-c_W)^{\sum_i (1-T_i) w_i + \beta_c - 1}$$

Updating censored values

In some surveys, high score values (i.e., S/CO) were censored. At each iteration, these censored values are imputed a S/CO from the corresponding score's conditional distribution.

For example, in a type 3 survey, we remind that the scores $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})^\top \sim N(\mu_s, \Sigma_s)$. If the first score - that is, X_{i1} of sample i was censored, then we impute a value - greater than the censored value for that score based on the distribution of X_{i1} conditional on (X_{i2}, X_{i3}, X_{i4}). That conditional distribution is easily obtained. Indeed, if U and V are two vectors (block variables) such that:

$$(U,V) \sim N\left(\mu = \left(\mu = (\mu_U, \mu_V)^\top, \Sigma = \begin{pmatrix} \Sigma_U & \Sigma'_{UV} \\ \Sigma_{UV} & \Sigma_V \end{pmatrix}\right)\right)$$
$$then\ U|V \sim N\left(\mu_U + \Sigma'_{UV}\Sigma_V^{-1}(V - \mu_V), \Sigma = \Sigma_U - \Sigma'_{UV}\Sigma_V^{-1}\Sigma_{UV}\right)$$

The conditional distribution of X_{i1} given (X_{i2}, X_{i3}, X_{i4}) is a particular case of the above (where $U = X_{i1}$ is of length 1) and can thus be easily calculated. The conditional distribution of X_{i2}, X_{i3} or X_{i4} given the other X_{i*} scores can be obtained along the same lines. We first left-multiply the vector X_i by the appropriate permutation matrix P to make the component X_{ij} of interest become the first component of $X_i^* = PX_i$, from which it follows that $X_i^* \sim N(P\mu, P'\Sigma P)$ and then obtain the conditional distribution for X_{ij} along the lines elicited above.