Statistics in Biosciences manuscript No.
(will be inserted by the editor)

Comparison of HIV prevalence among antenatal clinic attendees estimated from routine testing and unlinked anonymous testing

Ben Sheng · Jeffrey W Eaton · Mary Mahy · Le Bao

Received: March 11, 2019 / Accepted: date

Abstract In 2015, WHO and UNAIDS released new guidance recommending that countries transition from conducting antenatal clinic (ANC) unlinked anonymous testing (ANC-UAT) for tracking HIV prevalence trends among pregnant women to using ANC routine testing (ANC-RT) data, which are more consistent and economic to collect. This transition could pose challenges for distinguishing whether changes in observed prevalence are due to a change in underlying population prevalence or due to a change in the testing approach. We compared the HIV prevalence measured from ANC-UAT and ANC-RT in 15 countries that had both data sources in overlapping years. We used linear mixed-effects model (LMM) to estimate the RT-to-UAT calibration parameter as well as other unobserved quantities. We summarized the results at different levels of aggregation (e.g., country, urban, rural, and province). Based on our analysis, the HIV prevalence measured by ANC-UAT and ANC-RT data are consistent in most countries. Therefore, if large discrepancy is observed between ANC-UAT and ANC-RT at the same location, we recommend that people should be cautious and investigate the reason. For countries that lack information to estimate the calibration parameter, we propose an informative prior distribution of mean 0 and standard deviation 0.2 for the RT-to-UAT calibration parameter.

Ben Sheng
Department of Statistics, Pennsylvania State University
325 Thomas Building, University Park, PA 16802, USA
E-mail: bxs416@psu.edu

Jeffrey W Eaton
Department of Infectious Disease Epidemiology, Imperial College London
St Mary's Campus, Norfolk Place, London W2 1PG, United Kingdom
E-mail: jeffrey.eaton@imperial.ac.uk

Mary Mahy
UNAIDS
Avenue Appia 20, 1211 Geneve, Switzerland
E-mail: mahym@unaids.org

Le Bao
Department of Statistics, Pennsylvania State University
325 Thomas Building, University Park, PA 16802, USA
E-mail: lebao@psu.edu

Keywords Antenatal clinic · HIV prevalence · Unlinked anonymous HIV testing · Routine HIV testing · Calibration parameter · Linear mixed-effects model

1 Introduction

Monitoring trends in HIV prevalence is the cornerstone of HIV surveillance. Monitoring prevalence allows countries to estimate incidence trends and track progress toward reducing new HIV infections and ending the AIDS epidemic. To monitor trends in HIV prevalence, nearly all high HIV prevalence countries established HIV sentinel surveillance surveys at antenatal clinics (ANC) in the early 1990s, and repeated those surveys every one or two years. HIV prevalence data were collected from sentinel ANC sites chosen to provide a large enough sample size to monitor trends over time. The surveillance was based on unlinked anonymous testing (UAT) of women attending these selected ANCs (UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2003)).

In recent years, many countries have transitioned from conducting UAT sentinel surveillance surveys in favor of more efficient use of the routine testing (RT) done at ANCs (Dee et al. (2017)), and the data in the Spectrum files is evidence of this. The Spectrum software (Avenir Health, Glastonbury, CT, USA) provides trend estimates and values of key HIV indicators, and is used by UNAIDS and national programs for 161 countries. Within Spectrum, the Estimation and Projection Package (EPP) provides estimates of population HIV prevalence and incidence trends (Bao and Raftery (2010); Bao (2012); Bao et al. (2012); Brown et al. (2006); Brown et al. (2010); Brown et al. (2014); Ghys et al. (2004); Hogan and Salomon (2012); Sheng et al. (2017); Stover et al. (2012)). Since most high prevalence countries test all women attending ANC, the data from these routine tests were available to monitor the epidemic. There were numerous benefits in shifting to RT data for surveillance (Gourlay et al. (2015)). The use of UAT meant that women were not given the results of those tests (they were being tested a second time and those results were provided to them), and because the tests were anonymous, women were not given the option to opt out of the surveillance effort. On the other hand, for RT, the women receive their testing results, and the tests are voluntary. Many of the ANC-UAT sentinel sites had been selected because they were high volume sites, and were not representative of pregnant women in the country. In contrast, RT data includes all sites with available data, and generally does not have this site selection bias. Also, the number of women tested at routine ANC are much larger than the UAT sentinel surveillance, reducing sampling error and improving geographic specificity of the results. Finally, the financial and personnel resources previously spent on UAT sentinel surveys could now be directed to improve RT data.

Although both UAT and RT data allow countries to monitor the HIV prevalence among pregnant women, there could be systematic bias between them. The laboratory procedures for the UAT sentinel surveillance had more critical review due to careful protocol development and oversight by outside funders; as a result, the testing procedures might be more rigorous (Gonese et al. (2016)). During the course of UAT sentinel surveillance, the sites were provided testing kits to avoid stock outs. In RT, there might be stockouts during which health providers test women at increased risk of HIV infection, biasing prevalence estimates. Other biases might occur based on the management of the data from RT versus the lower volume and more rigorous UAT sentinel surveillance systems (Sirengo et al. (2016)). Therefore, it is of great interest to understand how the estimates of HIV prevalence would differ between the two data sources at different levels of aggregation (e.g., country, urban, rural, and province).

We also note that the switch from ANC-UAT to ANC-RT data leads to an extremely imbalanced sampling design; ANC-UAT were replaced by ANC-RT data around 2013. It challenges the estimation of HIV epidemics especially during the transition period. For instance, it could be difficult to distinguish

whether a change in observed prevalence is due to a change in underlying population prevalence or due to the systematic difference between two data sources. Therefore, the guideline encourages countries to explore the difference between ANC-UAT and ANC-RT data. In this manuscript, we focus on countries where both ANC-UAT and ANC-RT data are available with overlapping years, and aim to quantify the difference of HIV prevalence estimates inferred by those two surveillance systems.

2 Methods

We investigated countries in which both ANC-UAT and ANC-RT data were available from the same sites in the same years. We used linear mixed-effects model (LMM) to estimate the systematic bias in HIV prevalence from ANC-RT compared to ANC-UAT.

2.1 Prevalence Data from Antenatal Clinic (ANC) Sites

Several countries have collected UAT and RT data in the same years and at the same antenatal clinic (ANC) sites. Such datasets provide the opportunity to directly compare HIV prevalence rates estimated by the two data sources. For each country region, we only keep sites with "overlap" (UAT and RT in the same year) for analysis.

For this paper, we use the ANC prevalences given in the Spectrum files, and we explain the calculation procedures. (These prevalences were already calculated when we received the data, so we did not actually perform these calculations for our analyses. Nonetheless, we decide to provide the explanation.) To compute prevalence for ANC-UAT data, the number of pregnant women who tested positive for HIV was divided by the total number who had an HIV test (UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2003)). To compute prevalence for ANC-RT data, there needed to be data on (1) the number of pregnant women who already know their HIV positive status; (2) the number who had an HIV test, and (3) the number who tested positive for HIV. (In some cases, there may have been data on the number of women who are "previously known HIV-negative"; if this happened, the ANC-RT prevalence computation was changed accordingly.) (1) does rely on the accurate self-reporting of HIV status; however, excluding (1) would bias the ANC-RT prevalence estimate. ANC-RT prevalence was computed as (Sheng et al. (2017)):

$$\frac{[(1)\_known\_HIV\_positive] + [(3)\_tested\_HIV\_positive]}{[(1)\_known\_HIV\_positive] + [(2)\_total\_HIV\_tested]} \tag{1}$$

For ESA (Eastern and Southern Africa), there are 7 countries that have subnational areas with overlapping sites, including Ethiopia (Rural, Urban, 4 geographic regions), Lesotho (Urban), Namibia (14 geographic regions), Rwanda (Rural and Urban), Tanzania (17 geographic regions in total), Uganda (Rural and Urban), and Zimbabwe (7 geographic regions). For WCA (Western and Central Africa), there are 8 countries that have regions with overlapping sites, including Benin (12 geographic regions), Burkina Faso (Rural and Urban), Burundi (Rural and Urban), Cameroon (Rural and Urban), Ghana (Rural and Urban), Guinea Bissau (Rural and Urban), Senegal (1 general region), and Togo (Rural/Urban or 6 geographic regions). Togo sites are divided in two ways: 6 geographic regions (Centrale, Kara, Lome, Maritime, Plateaux, Savanes), and Rural/Urban split. Appendix A5 presents the data availability for each country region. The HIV epidemics are estimated at either the national level or the regional level (Brown et al., 2010, 2014; Stover et al., 2012). Therefore, we focus on investigating the difference between UAT prevalence and RT prevalence at the national level and the regional level.

2.2 Linear Mixed-Effects Model (LMM)

We adjust HIV prevalence from ANC data by $\frac{[\#\_of\_HIV\_positive]+0.5}{[sample\_size]+1}$; this adjustment prevents negative infinities at the probit scale, and enables the probit-transformed prevalences to follow an approximately Normal distribution. Let W_{st}^{UAT} and W_{st}^{RT} be ANC-UAT and ANC-RT observed prevalences at probit scale for Site s and Year t, and ρ_t be the true UAT prevalence at probit scale in Year t. We use the probit transformation, so it is consistent with the transformation used in previous models for ANC prevalence (Alkema et al. (2007); Brown et al. (2014); Sheng et al. (2017)). The RT calibration parameter β represents the mean probit difference between UAT prevalence and RT prevalence, and is the main parameter of interest. b_s is the site effect for Site s, modeled as a random effect with variance ξ^2. ϵ_{st}^{UAT} and ϵ_{st}^{RT} are random errors for UAT and RT, and σ^2 and τ^2 are the corresponding UAT and RT residual variance parameters. We use the following model to obtain the national estimates of β.

$$W_{st}^{UAT} = \rho_t + b_s + \epsilon_{st}^{UAT} \tag{2}$$

$$\epsilon_{st}^{UAT} \sim N(0, \sigma^2) \tag{3}$$

$$W_{st}^{RT} = \beta + \rho_t + b_s + \epsilon_{st}^{RT} \tag{4}$$

$$\epsilon_{st}^{RT} \sim N(0, \tau^2) \tag{5}$$

$$b_s \sim N(0, \xi^2) \tag{6}$$

To obtain sub-national estimates of β, we apply the stratified model below to countries with multiple regions (all countries except Lesotho and Senegal), where i stands for a particular region. The stratified model adds region-level fixed effects for RT calibration β_i and UAT true prevalence ρ_{it}. For both the country-level and stratified region models, we fit the LMM using the "nlme" R package, and present the runtimes in Appendix A6.

$$W_{ist}^{UAT} = \rho_{it} + b_s + \epsilon_{ist}^{UAT} \tag{7}$$

$$\epsilon_{ist}^{UAT} \sim N(0, \sigma^2) \tag{8}$$

$$W_{ist}^{RT} = \beta_i + \rho_{it} + b_s + \epsilon_{ist}^{RT} \tag{9}$$

$$\epsilon_{ist}^{RT} \sim N(0, \tau^2) \tag{10}$$

The model assumes that site-level probit scale prevalence trends only differ by a site effect b_s, which is a constant over time. This assumption might be too strong over a long time period because different sites might experience different rates of prevalence change (e.g., sites with higher prevalence might have greater prevalence declines). Therefore, we truncate the early years of historical UAT data, and only keep the UAT data years that overlap with RT data years; it leads to a range of 3-8 years of data. See Appendix A5 for more details. We also run the full data analysis without truncating the early years. If the results between truncated data analysis and full data analysis are significantly different, we would recommend using the truncated data results because parameters are more likely to be constant over a shorter time period.

In addition to constant site effect, the model also assumes a constant RT calibration parameter; we discuss this assumption in Section 2.3. We assume conditional independence between the UAT residuals (ϵ_{st}^{UAT} or ϵ_{ist}^{UAT}) and the RT residuals (ϵ_{st}^{RT} or ϵ_{ist}^{RT}); this assumption is discussed in Section 2.4.

2.3 Assumption of Constant RT Calibration

In this section, we investigate the assumption of constant RT calibration parameter β on truncated data. For countries (and country regions) with multiple years of overlap between UAT and RT, we split the truncated data years into two parts (each part must contain both UAT data and RT data), and fit the following model that allows the RT calibration parameter to differ between two parts.

$$W_{st}^{UAT} = \rho_t + b_s + \epsilon_{st}^{UAT} \tag{11}$$
$$\epsilon_{st}^{UAT} \sim N(0, \sigma^2) \tag{12}$$
$$W_{st}^{RT} = \beta_{\{t \le T\}} + \beta_{\{t > T\}} + \rho_t + b_s + \epsilon_{st}^{RT} \tag{13}$$
$$\epsilon_{st}^{RT} \sim N(0, \tau^2) \tag{14}$$
$$b_s \sim N(0, \xi^2) \tag{15}$$

As a comparison, we also fit the "base" model that assumes the constant RT calibration:

$$W_{st}^{RT} = \beta + \rho_t + b_s + \epsilon_{st}^{RT} \tag{16}$$

For each country (and country region), we use BIC to select the best model among all possible "split" models and the "base" model. We can estimate the "split" RT model in 49 cases with multiple years of overlap. In 44 out of 49 cases, the "base" model has lowest BIC; in the remaining cases, the maximum difference between $\beta_{\{t \le T\}}$ and $\beta_{\{t > T\}}$ are less than 0.5. Thus, the assumption of constant RT calibration parameter is reasonable for most truncated cases. We present the detailed results in Appendix A3.

2.4 Assumption of Independence between UAT and RT Residuals

Since UAT is anonymous, there is no empirical evidence about whether the same woman may participate in both UAT and RT. However, it can potentially introduce correlation between UAT and RT if a large portion of women are involved in both UAT and RT testing. In this section, we examine the assumption of conditional independence between UAT residuals (ϵ_{st}^{UAT} or ϵ_{ist}^{UAT}) and RT residuals (ϵ_{st}^{RT} or ϵ_{ist}^{RT}) on truncated data.

For each country (and country region), on truncated data, we fit the "correlated" model below, and test whether the UAT/RT correlation parameter ρ significantly differs from 0.

$$W_{st}^{UAT} = \rho_t + b_s + \epsilon_{st}^{UAT} \tag{17}$$
$$\epsilon_{st}^{UAT} \sim N(0, \sigma^2) \tag{18}$$
$$W_{st}^{RT} = \beta + \rho_t + b_s + \epsilon_{st}^{RT} \tag{19}$$
$$\epsilon_{st}^{RT} \sim N(0, \tau^2) \tag{20}$$
$$Corr(\epsilon_{st}^{UAT}, \epsilon_{st}^{RT}) = \rho \tag{21}$$
$$b_s \sim N(0, \xi^2) \tag{22}$$

Note that we do not fit a stratified region model for this analysis. Instead, we subset the data by region to avoid information sharing across regions when computing correlation.

We present the results in Appendix A4. There is sufficient data to estimate the UAT/RT correlation parameter and its standard error in 72 cases, and in most cases (59 out of 72), there is no significant UAT/RT correlation at significance level 0.05. Furthermore, the "correlated" models and the corresponding "independent" models produce similar point estimates of the RT calibration parameter. Thus, we do not think this potential correlation will be an issue when estimating RT calibration on truncated data.

2.5 Maximum Likelihood Estimation (MLE)

As an alternative approach to LMM, we use maximum likelihood estimation (MLE) to estimate model parameters, and present the MLE results in Appendix A2. The difference in statistical model is the site-effect; for MLE, we replace the random effects of LMM with centered fixed effects, where M is the number of sites.

$$\sum_{s=1}^{M} b_s = 0 \tag{23}$$

We derive closed formed solutions for the maximum likelihood estimators, and show the derivations in Appendix A2.

$$\hat{\beta} = \frac{1}{|RT|} \sum_{(s,t)\in RT} (W_{st}^{RT} - \hat{\rho}_t - \hat{b_s}) \tag{24}$$

$$\hat{\sigma^2} = \frac{1}{|UAT|} \sum_{(s,t)\in UAT} (W_{st}^{UAT} - \hat{\rho}_t - \hat{b_s})^2 \tag{25}$$

$$\hat{\tau^2} = \frac{1}{|RT|} \sum_{(s,t)\in RT} (W_{st}^{RT} - \hat{\rho}_t - \hat{b_s} - \hat{\beta})^2 \tag{26}$$

$$\hat{\rho}_t = \frac{I(t \in UAT) * \hat{\tau^2} * \sum_{s\in UAT_t} (W_{st}^{UAT} - \hat{b_s}) + I(t \in RT) * \hat{\sigma^2} * \sum_{s\in RT_t} (W_{st}^{RT} - \hat{b_s} - \hat{\beta})}{\hat{\tau^2} * |UAT_t| + \hat{\sigma^2} * |RT_t|} \tag{27}$$

For $s = 1, ..., M-1$,

$$\hat{b_s} = \frac{\hat{\tau^2} * A_s + \hat{\sigma^2} * B_s}{\hat{\tau^2} * (|UAT_s| + |UAT_M|) + \hat{\sigma^2} * (|RT_s| + |RT_M|)} \tag{28}$$

$$A_s = \sum_{t\in UAT_s} (W_{st}^{UAT} - \hat{\rho}_t) - \sum_{t\in UAT_M} (W_{Mt}^{UAT} - \hat{\rho}_t) - |UAT_M| * \sum_{i\notin(s,M)} \hat{b_i} \tag{29}$$

$$B_s = \sum_{t\in RT_s} (W_{st}^{RT} - \hat{\rho}_t - \hat{\beta}) - \sum_{t\in RT_M} (W_{Mt}^{RT} - \hat{\rho}_t - \hat{\beta}) - |RT_M| * \sum_{i\notin(s,M)} \hat{b_i} \tag{30}$$

For each country (or country region), we subset the data by country (or country region), and compute the MLE in the following iterative fashion.

1. Set initial values for $\hat{\rho}_t$ and $\hat{b_s}$.
   (a) $\hat{\rho}_t = \frac{1}{|UAT_t|} \sum_{s\in UAT_t} W_{st}^{UAT}$ for each Year t.
   (b) $\hat{b_s} = 0$ for each Site s.
2. Compute $\hat{\beta}$.
3. Compute $\hat{\sigma^2}, \hat{\tau^2}$.
4. Compute $\hat{\rho}_t$ for each Year t.
5. Compute $\hat{b_s}$ for each Site s.
   (a) Compute $\hat{b_1}, ..., \hat{b_{M-1}}$ using the MLE formula.
   (b) Compute $\hat{b_M} = -\sum_{s=1}^{M-1} \hat{b_s}$.
   (c) Repeat (a) and (b) until $\hat{b_s}$ changes less than .001 for 10 consecutive iterations for all sites.
6. Repeat Steps 2-5 until $\hat{\beta}$ changes less than .001 for 10 consecutive iterations.

To obtain standard errors (SE) of model parameters, we compute the information matrix based on the Hessian matrix, and then compute the inverse of the information matrix. The diagonal of the inverse matrix represents the variance of each parameter, and its square root is the SE (Lehmann (2004)).

As shown in the aggregation summaries, the main conclusions between MLE and LMM are essentially the same. However, the LMM results include more country regions; in particular, 6 regions in Tanzania (Arusha, Dodmoa, Kagera, Morogoro, Mwanza, Tabora) could be fit using LMM but not MLE. Also, LMM includes standard error for all countries and country regions, while MLE sometimes encounters problems with the information matrix.

3 Results

In this section, we first present the complete results for Uganda Rural; this includes the estimates for RT calibration, UAT prevalence, variances (UAT residual, RT residual, site effect), and site effects. Then, the estimates of RT calibration are presented for 7 countries in ESA (Eastern and Southern Africa) and 8 countries in WCA (Western and Central Africa).

The remaining results are not presented in this section. Appendix A7 presents estimated variances (UAT residual, RT residual, site effect) for country-level and shared stratified. We also estimated site-effects at country-level and shared stratified, and UAT prevalence for each country and country region; however, we do not present these.

3.1 Complete Results for Uganda Rural

This subsection visualizes results for a particular stratified country region. In the stratified model, the RT calibration and UAT prevalences are estimated at the regional level, while the variance parameters and site effects are estimated at the country level. We present the results based on full data (1989-2017 for Uganda Rural and 1986-2017 for Uganda country) and truncated data (2012-2017 for Uganda).

For RT calibration of Uganda Rural, full data gives estimate 0.009 and standard error (SE) 0.044, while truncated data gives estimate 0.007 and SE 0.039; full data and truncated data estimates are not significantly different. For UAT prevalence of Uganda Rural, we first compute the 95% confidence intervals (CI) (Estimate $\pm$ 1.96*SE) on the probit scale, and then convert to the probability scale. The UAT prevalence estimates from full and truncated data also appear to be similar according to Figure 1.

The stratified variance parameters (UAT residual variance σ^2, RT residual variance τ^2, and site effect variance ξ^2) are shared across all regions within a particular country, and Table 1 presents the estimates for Uganda. The estimated variance parameters are also not significantly different between full data and truncated data. The estimated Uganda site effects are mostly similar between the truncated and full data analyses, and they are presented in Figure 2 (sorted by site effects for truncated data).

Parameter	**Full Data**		**Truncated Data**	
	Estimate	*95% CI*	*Estimate*	*95% CI*
σ^2	0.045	(0.038,0.054)	0.036	(0.024,0.055)
τ^2	0.026	(0.020,0.033)	0.020	(0.016,0.026)
ξ^2	0.050	(0.028,0.088)	0.057	(0.032,0.100)

Table 1: Estimated shared variance parameters for Uganda regions.

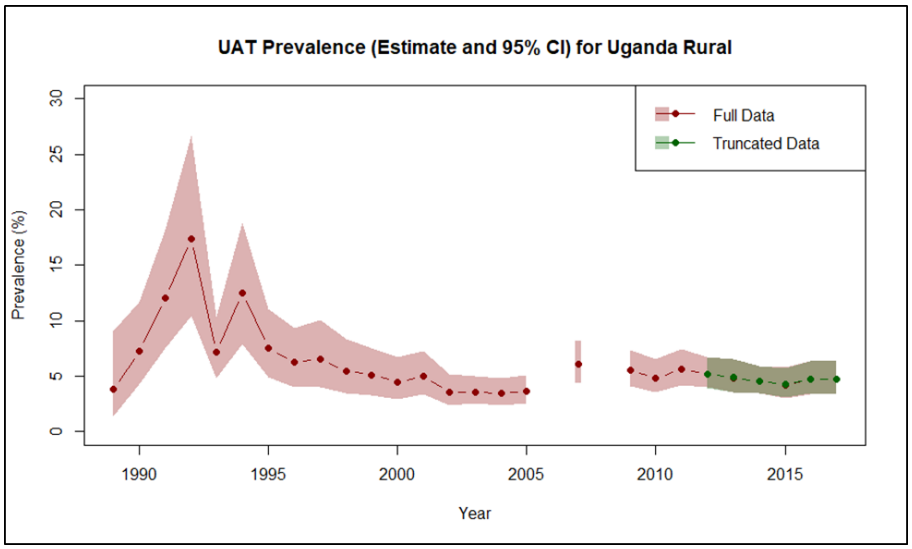

Fig. 1: Estimated UAT prevalence for Uganda Rural.

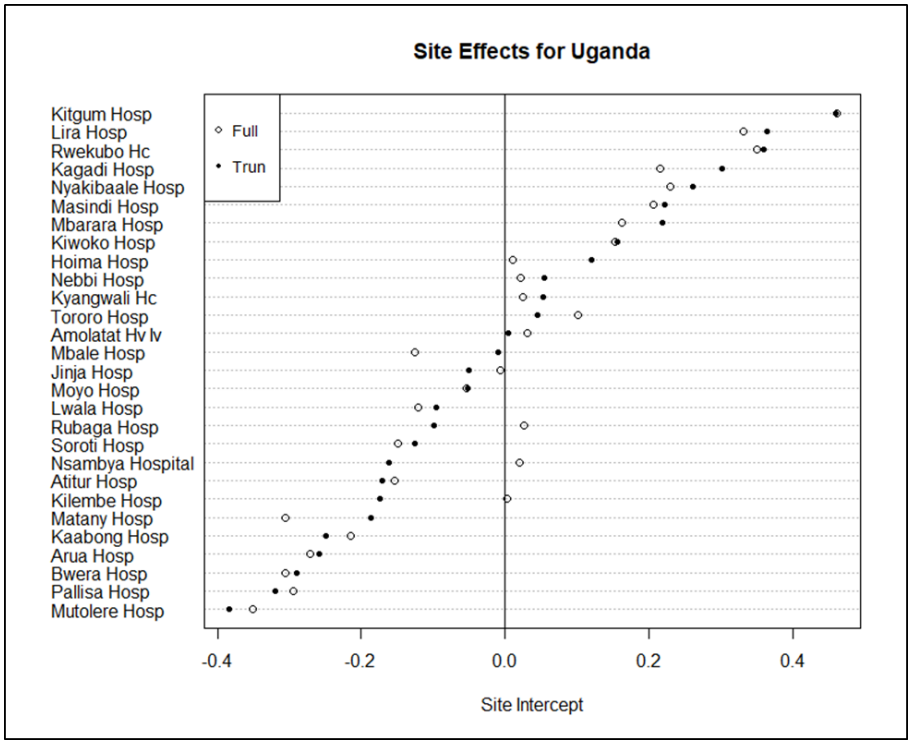

Fig. 2: Estimated Uganda site effects (sorted by site effects for truncated data).

3.2 RT Calibration for 15 African Countries

In the tables and figures of this section, we present the RT calibration results for 7 countries in ESA (Eastern and Southern Africa), including Ethiopia, Lesotho, Namibia, Rwanda, Tanzania, Uganda, and Zimbabwe, and 8 countries in WCA (Western and Central Africa), including Benin, Burkina Faso, Burundi, Cameroon, Ghana, Guinea Bissau, Senegal, and Togo.

We present results by full/truncated data. In general, we believe that the truncated analysis is more robust than the full analysis because we only need to assume the RT calibration and site effects are constant in a short time window. Nonetheless, as the figures and tables of this section show, the RT calibration results are similar from full data and truncated data.

Table 2 presents the RT calibration results for countries, by full/truncated data. The full data analysis could have a higher SE than the truncated data analysis; while the full data analysis has more data points, it also involves more parameters (from additional prevalence years), which might increase SE. We present the stratified country region results in Appendix A1.

Global Area	**Country**	**Full Data**		**Truncated Data**	
		Estimate	*SE*	*Estimate*	*SE*
Eastern and Southern Africa	*Ethiopia*	-0.084	0.026	-0.090	0.027
	Lesotho	0.107	0.061	0.107	0.039
	Namibia	-0.026	0.022	-0.027	0.021
	Rwanda	0.016	0.027	0.016	0.023
	Tanzania	-0.084	0.043	-0.082	0.040
	Uganda	0.046	0.036	0.041	0.032
	Zimbabwe	-0.063	0.030	-0.063	0.019
Western and Central Africa	*Benin*	-0.125	0.022	-0.125	0.022
	Burkina Faso	-0.084	0.030	-0.083	0.038
	Burundi	-0.214	0.164	-0.214	0.174
	Cameroon	0.147	0.059	0.147	0.063
	Ghana	-0.154	0.028	-0.154	0.027
	Guinea Bissau	-0.041	0.056	-0.044	0.058
	Senegal	-0.237	0.073	-0.237	0.077
	Togo	-0.118	0.066	-0.118	0.060

Table 2: RT calibration results at country level.

Figure 3 presents the RT calibration estimate with 95% CI for the total country files of ESA and WCA. The full and truncated data analyses have overlapping 95% CI for all countries, suggesting no significant difference. We present the stratified country region results in Appendix A1.

3.3 Summary of RT Calibration Estimates

Table 3 summarizes the point estimates of RT calibration parameters from Section 3.2 and Appendix A1, by level of aggregation and by full/truncated data. From Table 3, we conclude that the aggregation means are not significantly different from 0. Also, the unadjusted standard deviations (SD) (assuming means listed in Table 3) are all less than .200, and the adjusted standard deviations (adj-SD) (assuming mean 0) are less than .205.

Senegal has only 1 general region, so it is counted in "Country" but not in the other aggregation levels. Lesotho has only 1 region with overlapping UAT/RT (Urban), so it is counted in both "Country"

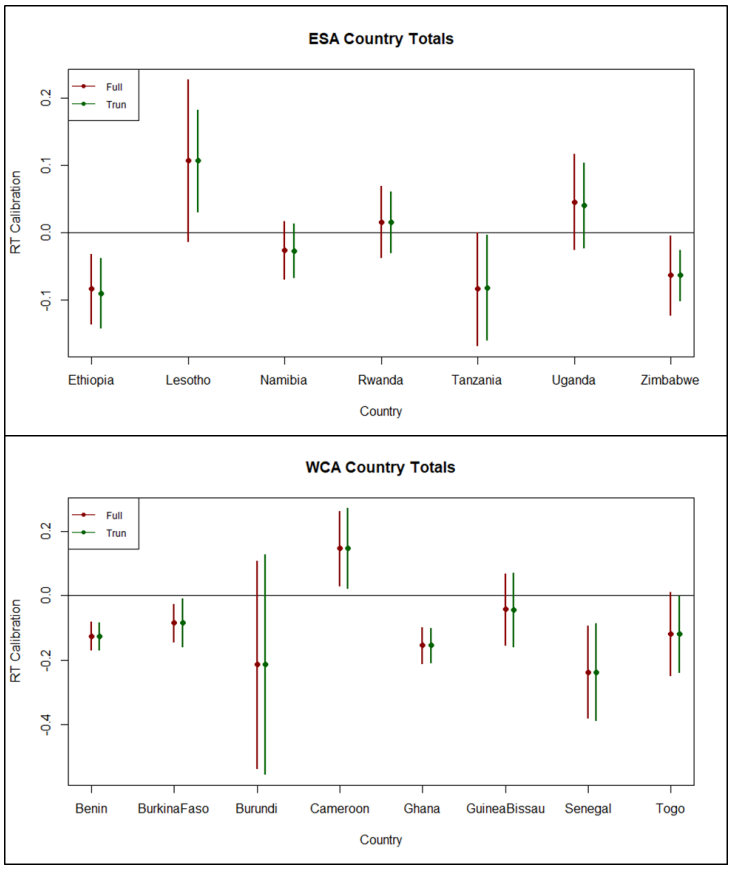

Fig. 3: RT calibration estimates for ESA and WCA total country files.

and "Urban" aggregations. However, since Lesotho Total and Lesotho Urban are the same, it only counts once in "All Levels"; thus, the count for "All Levels" is 1 less than the sum of the counts from "Country", "Rural", "Urban", and "Province".

4 Discussion

We compared observed HIV prevalence data from unlinked anonymous testing (UAT) and routine testing (RT), and assessed the difference between UAT and RT at different levels of aggregation. It leads to an informative prior distribution for the calibration parameter between UAT and RT for models that would jointly use both data sources.

For instance, the Estimation and Projection Package (EPP) has been used by many countries to generate HIV epidemic estimates that determine their policy and resource allocation. Currently, the EPP model uses a standard normal distribution for the RT calibration parameter (Sheng et al. (2017)). Based on our analysis, the mean of the RT calibration parameter should remain at 0. However, our results suggest the standard deviation could be reduced from 1.0 to 0.2, which suggests the ANC-RT prevalence

Level of Aggregation	Statistic	Full Data	Truncated Data
All Levels	Count	93	93
	Mean	-0.079	-0.080
	SD	0.134	0.133
	adj-SD	0.156	0.156
	Min	-0.540	-0.540
	Max	0.403	0.403
Country	Count	15	15
	Mean	-0.061	-0.062
	SD	0.108	0.107
	adj-SD	0.125	0.125
	Min	-0.237	-0.237
	Max	0.147	0.147
Rural	Count	9	9
	Mean	-0.030	-0.031
	SD	0.133	0.133
	adj-SD	0.137	0.137
	Min	-0.285	-0.285
	Max	0.151	0.151
Urban	Count	10	10
	Mean	-0.059	-0.061
	SD	0.193	0.191
	adj-SD	0.202	0.201
	Min	-0.426	-0.426
	Max	0.197	0.197
Province	Count	60	60
	Mean	-0.091	-0.091
	SD	0.131	0.130
	adj-SD	0.160	0.160
	Min	-0.540	-0.540
	Max	0.403	0.403

Table 3: Summary of RT calibration point estimates by level of aggregation.

is more consistent with ANC-UAT prevalence than encapsulated by the current non-informative prior distribution.

We used a relatively simple approach to investigate the difference between ANC-UAT and ANC-RT data. We did not use the EPP itself to produce estimates in our analysis; EPP is built on a susceptible infected model, which imposes a certain structure on the epidemic trends and could interact with the estimate of the RT calibration parameter (Bao and Raftery (2010); Bao (2012); Bao et al. (2012); Brown et al. (2006); Brown et al. (2010); Brown et al. (2014); Ghys et al. (2004); Hogan and Salomon (2012); Sheng et al. (2017); Stover et al. (2012)). Instead, we decided to use the linear mixed-effects model, for parameter estimation; one benefit is the computation speed, as it can fit all regions in a country (full data or truncated data) in less than 10 seconds.

The constant site effect assumption has less impact in our analysis based on truncated years because we only need to assume the site effects are constant in a short time window. Therefore, we decide to draw the final conclusion on RT calibration by using the truncated years analysis. However, as the figures and tables show, the RT calibration results are similar from full data and truncated data.

In this analysis, we assume that each woman will only be tested once per year. For UAT, this is the case. For RT, the guidance is that only the first test during a pregnancy for each woman should be reported. There is concern that this is not (able to be) followed, and they report all tests. This is potentially a source of bias because HIV positive women would not need to be tested again, but the

guidelines are that HIV negative women are tested again later during pregnancy, so selectively re-testing HIV negative women. In the future, we could investigate the case where multiple tests per year are reported for some women.

The model utilizes several conditional independence assumptions. Section 2.4 tested the independence of UAT residuals and RT residuals. Conditional independence is also assumed across sites for UAT residuals, RT residuals, and site effects (b_s); we believe relatively few women would be tested at multiple sites. Also, we assume conditional independence across years for UAT residuals and RT residuals. Our main conclusion is based on the truncated data analysis, and we do not expect a large fraction of women to have done multiple tests across 3-8 years (also, some HIV negative women would become HIV positive over time, which would reduce potential time correlation); thus, we do not believe that correlation across years would be large enough to change our conclusion. Finally, we assume conditional independence between UAT/RT residuals and the site effects. While we believe these assumptions should not be a major problem in our analysis, future studies could test the conditional independence assumptions across sites, across years, and between UAT/RT residuals and site effects.

For future work, we could incorporate the proposed informative prior distribution into EPP, and evaluate the effect on the estimates and also the efficiency of convergence. Also, we should continually update the analysis using newly collected data, as there could be additional years of overlap within regions already analyzed, and regions previously without overlap might now have overlap.

Acknowledgements
This research was supported by the Joint United Nations Programme on HIV/AIDS (UNAIDS), NIH/NIAID 5-R01-AI136664, NIH 5-R03-AI125001, NSF IGERT - DGE-1144860. The authors are also grateful to Kimberly Marsh and Kaiyi Wu.

References

Alkema L, Raftery AE, Clark SJ (2007) Probabilistic projections of HIV prevalence using Bayesian melding. The Annals of Applied Statistics 1(1):229–248

Bao L (2012) A new infectious disease model for estimating and projecting HIV/AIDS epidemics. Sex Transm Infect 88(Suppl 2):i58–i64

Bao L, Raftery AE (2010) A stochastic infection rate model for estimating and projecting national HIV prevalence rates. Sexually Transmitted Infections 86(Suppl 2):ii93–ii99

Bao L, Salomon JA, Brown T, Raftery AE, Hogan DR (2012) Modelling national HIV/AIDS epidemics: revised approach in the UNAIDS Estimation and Projection Package 2011. Sex Transm Infect 88(Suppl 2):i3–i10

Brown T, Grassly NC, Garnett G, Stanecki K (2006) Improving projections at the country level: the UNAIDS Estimation and Projection Package 2005. Sexually Transmitted Infections 82(suppl 3):iii34–iii40

Brown T, Bao L, Raftery AE, Salomon JA, Baggaley RF, Stover J, Gerland P (2010) Modelling HIV epidemics in the antiretroviral era: the UNAIDS Estimation and Projection package 2009. Sexually Transmitted Infections 86(Suppl 2):ii3–ii10

Brown T, Bao L, Eaton JW, Hogan DR, Mahy M, Marsh K, Mathers BM, Puckett R (2014) Improvements in prevalence trend fitting and incidence estimation in EPP 2013. AIDS (London, England) 28(4):S415

Dee J, Calleja JMG, Marsh K, Zaidi I, Murrill C, Swaminathan M (2017) HIV surveillance among pregnant women attending antenatal clinics: evolution and current direction. JMIR Public Health and Surveillance 3(4)

Ghys PD, Brown T, Grassly NC, Garnett G, Stanecki KA, Stover J, Walker N (2004) The UNAIDS Estimation and Projection Package: a software package to estimate and project national HIV epidemics. Sexually Transmitted Infections 80(suppl 1):i5–i9

Gonese E, Mushavi A, Mungati M, Mhangara M, Dzangare J, Mugurungi O, Dee J, Kilmarx PH, Shambira G, Tshimanga MT, Hargrove J (2016) Is Zimbabwe ready to transition from anonymous unlinked sero-surveillance to using prevention of mother to child transmission of HIV (PMTCT) program data for HIV surveillance?: results of PMTCT utility study, 2012. BMC Infectious Diseases 16(1):97

Gourlay A, Wringe A, Todd J, Michael D, Reniers G, Urassa M, Njau P, Kajoka D, Lema L, Zaba B (2015) Challenges with routine data sources for PMTCT programme monitoring in East Africa: insights from Tanzania. Global Health Action 8(1):29987

Hogan DR, Salomon JA (2012) Spline-based modelling of trends in the force of HIV infection, with application to the UNAIDS Estimation and Projection Package. Sex Transm Infect 88(Suppl 2):i52–i57

Lehmann EL (2004) Elements of large-sample theory. Springer Science & Business Media, New York, NY, USA

Sheng B, Marsh K, Slavkovic AB, Gregson S, Eaton JW, Bao L (2017) Statistical models for incorporating data from routine HIV testing of pregnant women at antenatal clinics into HIV/AIDS epidemic estimates. AIDS (London, England) 31(Suppl 1):S87

Sirengo M, Rutherford GW, Otieno-Nyunya B, Kellogg TA, Kimanga D, Muraguri N, Umuro M, Mirjahangir J, Stein E, Ndisha M, Kim AA (2016) Evaluation of Kenya's readiness to transition from sentinel surveillance to routine HIV testing for antenatal clinic-based HIV surveillance. BMC Infectious Diseases 16(1):113

Stover J, Brown T, Marston M (2012) Updates to the Spectrum/Estimation and Projection Package (EPP) model to estimate HIV trends for adults and children. Sex Transm Infect 88(Suppl 2):i11–i16

UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2003) Guidelines for conducting HIV sentinel serosurveys among pregnant women and other groups. Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO), Geneva, Switzerland