1
2
3
4
5 **Potential Utility of Routine Programmatic Data in Monitoring National and State-**
6 **Level HIV Epidemic in Nigeria: Data Triangulation Analysis**
7
8
9

10
11 Debem Henry^{1,*}, Aminu Yakubu^{1,¶}, Mukhtar Ahmed^{1,¶}, Gwamna Jerry^{1,&}, Dalhatu Ibrahim^{1,&}
12

13
14
15 ¹Division of Global HIV, U.S. Centers for Disease Control and Prevention, Abuja, Nigeria
16
17

18
19 *Corresponding Author
20 Email: wqe4@cdc.gov (DH)
21

22 [&]These authors contributed equally to this work
23

24 [¶]These authors contributed equally to this work
25
26

27 **DISCLAIMER**

28 The findings and conclusions in this report are those of the authors and do not necessarily represent
29 the official position of the funding agencies. Use of trade names is for identification only and does not
30 imply endorsement by CDC or the United States Department of Health and Human Services.

Abstract

Nigeria relies on data from periodic resource-intensive surveys such as antenatal HIV seroprevalence sentinel surveys (ANC-HSS) and population-based National AIDS and Reproductive Health Surveys (NARHS) for its HIV control efforts. Nigeria has not explored the use of readily available routine programmatic data (RPD) to easily inform and monitor epidemic control efforts at local settings in near real time. This study aimed to determine the utility of RPDs (Prevention of Mother-To-Child Transmission [PMTCT] and HIV Testing and Counseling [HTC]) as a proxy for monitoring HIV epidemic in Nigeria. Using World Health Organization 12 step triangulation procedures, we compared state-level seropositivity data from PMTCT and HTC programs to HIV prevalence data from NARHS and ANC-HSS reports in relevant pairs from 2010 to 2014 in Nigeria. The study population was pregnant women and general population. We abstracted relevant data from PEPFAR Nigeria data source and published national survey reports. We compared visual (scatterplots and maps) patterns and trends, and performed Pearson correlation and univariate linear regression models of the estimates for best matched/contiguous years for which data were available. Correlation between PMTCT2014 and ANC-HSS2014 was positive and significant ($R=0.7, p<0.001$). ANC-HSS2014 and HTC2014 were slightly correlated ($R=0.4, p<0.05$). Significant correlation was observed between ANC-HSS2010 and PMTCT2013 ($R=0.8, p<0.001$) and between ANC-HSS2010 and HTC2013 ($R=0.6, p<0.001$). All RPD sources and ANC-HSS indicated a decreasing trend in national HIV prevalence in Nigeria. PMTCT2014 data showed strong capability of predicting HIV prevalence in ANC-HSS2014 in regression model ($B=2.09, p<0.0001$). Use of routine PMTCT data in monitoring HIV prevalence among women of reproductive age could be more valid and reliable in local settings than the use of HTC data. Use of RPD to monitor national and sub-national-level HIV epidemic in between national surveys in Nigeria could maximize program resources, and promote a more responsive and efficient actions toward epidemic control.

Introduction

An estimated 3.2 million people live with HIV (PLHIV) in Nigeria with 160,000 of those dying annually [1]. Additionally, the number of new infections was estimated at 220,000 in 2013 [1] which is the second highest in the African sub-region, while the disease has also rendered 1.8 million children orphans [2]. However, Nigeria has estimated a steady decline in the prevalence of HIV, from 5.8% in 2001, to 4.1% in 2010 and 3.0% in 2014 [3].

Common sources of HIV survey data are the Antenatal Care HIV Sero-Prevalence Sentinel Surveys (ANC-HSS) and the National AIDS and Reproductive Health Survey (NARHS). These surveys are designed for obtaining data for specific HIV related estimates that could be useful in population based health planning and programming. The ANC-HSS is conducted every 2 years [4] and NARHS every 4 years. ANC-HSS is conducted among all pregnant women who visit an antenatal facility for antenatal services for the first time for a confirmed pregnancy [5]. The survey was conducted in about 160 antenatal care (ANC) sentinel facilities across the 36 states and the Federal Capital Territory (FCT)[6]. The NARHS is a nationally representative household survey conducted to elicit information from the general population on relevant HIV, reproductive health, and knowledge and behavioral indicators [7]. Though the NARHS is considered to be more nationally representative, it is more expensive and difficult to sustain in resource-limited settings.

Prevention-of-Mother- to-Child Transmission (PMTCT) of HIV services involves the identification of pregnant women living with HIV during ANC and the provision of appropriate Antiretroviral Therapy (ART) to ensure that the mother stays healthy and prevent HIV transmission to the child [8]. The PMTCT services are provided in most ANC clinics so the women attending ANC services are easily referred to PMTCT services in the same clinic. Technically, routine PMTCT data and ANC-HSS both reflect HIV testing among pregnant women, however PMTCT data does not cover all ANC sites. Among pregnant women,

79 enrolment into ART is preceded by an opt-out HIV test and counseling (HTC) process that determines
80 the HIV status of the clients (including PMTCT clients). For the general population, HTC services are
81 usually the entry points into HIV programs in which the clients are referred to access other HIV services
82 (prevention, treatment, and care services) depending on their HIV status. Unlike the PMTCT services,
83 which is limited to pregnant women, the HTC services are open to the general population in the clinics
84 and communities through mobile community testing.

85 PMTCT data are primarily collected and generated at the HTC settings of the ANC clinics using
86 the national PMTCT registers. The data are summarized on monthly basis by each facility using the
87 Monthly Summary Forms (MSF) and transmitted manually or electronically to the next higher sub-
88 national unit reporting/data collection level for validation, collation, and further transmission to the
89 national/central level. HTC data, sharing similar data reporting system with the PMTCT, is generated
90 from all other HTC points of service operated at other settings within the clinic (such as voluntary
91 counselling and testing, DOT-tuberculosis, out-patient-department settings etc) and community
92 settings, except from the ANC; thus, its population is more generalized than the PMTCT and more likened
93 to the NARHS population that involved household (community) testing.

94 Researchers have recommended the need to validate both the use of routine programmatic data
95 (RPD) such as PMTCT program data for surveillance purposes [6,9–11] and the suitability of the ANC-HSS
96 among other data sources in estimating the HIV prevalence in the general population [12–15]. These
97 estimations form the essential principles of the Joint United Nations Programme on HIV/AIDS (UNAIDS)
98 spectrum HIV estimates for many countries [16]. Some countries such as Kenya, Ethiopia [6] and Rwanda
99 [10] have been able to integrate the PMTCT information in the ANC forms which makes it easier to collect
100 and compare information from both arms (RPD and Survey data) during the countries' ANC-HSS.

101 However, Nigeria has not been able to successfully link these two arms in ANC-HSS; hence, PMTCT data
102 utility for surveillance and monitoring purposes has not been fully established in the country.

103 **Rationale for the study**

104 The government of Nigeria (GoN) and PEPFAR have supported the generation and management
105 of routine program data primarily for accountability and the monitoring of the PMTCT program, as well
106 as to improve the implementation of HIV programs and services in Nigeria. However, the utility of
107 program data in monitoring the country's epidemic pattern and trend, and for timely impact assessment
108 of the HIV programs in Nigeria has not been previously demonstrated. In addition, producing reliable,
109 timely and consistent surveillance and population based data to support the effective implementation
110 of evidence-based high-impact interventions to control HIV epidemic remains a challenge. The aim of
111 the study was to determine the potential utility of RPD in monitoring HIV epidemic among the pregnant
112 women and general population at the national and local levels in Nigeria. The significance of the study
113 also shares in the benefits of effective HIV case base surveillance systems in which updated population
114 level data that are routinely collected from health records are used to inform epidemic control decisions
115 at both local and regional levels [17]; hence, accommodating the dynamism of the epidemic. The
116 overarching research question was: to what extent could routine PMTCT and HTC program data sources
117 be used to monitor HIV prevalence among women of reproductive age and in general population in the
118 sub-national settings in Nigeria? This activity was reviewed in accordance with the US Centers for Disease
119 Control and Prevention (CDC) human research protection procedures and was determined to be
120 research, but did not involve human subjects.

Objectives of the study

- I. To determine data concurrence and any correlation(s) between HIV seropositivity estimates from RPD (PMTCT and HTC data) and prevalence estimates from the national survey/surveillance data (ANC-HSS and NARHS) sources between 2010 and 2014 in Nigeria.
- II. To examine the trend and concurrence between the national HIV seropositivity rate from RPD and national prevalence from the national survey/surveillance data (NSD).
- III. To determine the extent of predictive association between the HIV seropositivity estimates from RPD and prevalence from the NSD sources.

Materials and methods

Study design

The target population were the general population and pregnant women. We conducted a retrospective secondary analysis of HIV seropositivity data from RPD sources and compared them with those reported in NSD sources: ANC-HSS (for pregnant women) and NARHS survey (for general population) reports. The comparison, in principle, adopted the 12- step triangulation model recommended by WHO [18]. The model essentially involves initial identification of research questions that are answerable through triangulation, then, identifying the data sources, understanding the background of the data sources, collation of the data/reports, running observational analysis to understand the pattern and trend among the sources, and drawing conclusions through summarized findings [18]. Data sources were identified based on the objectives of this study. The study variables from the data sources were the HIV prevalence estimates from NSD and seropositive rates from RPD.

Data Sources

142 The relevant data sources were broadly categorized into two: NSD and RPD sources. The NSDs
 143 were the ANC-HSS and NARHS survey reports while the RPDs were the PEPFAR HTC and PMTCT APR
 144 data (see Table 1A below). The variables of interest from these data sources were the HIV prevalence
 145 estimates (from ANC-HSS and NARHS) and seropositivity rates (from PMTCT and HTC). The ANC-HSS
 146 and PMTCT data sources represent prevalence estimates among pregnant women while the NARHS
 147 and HTC data sources represent that in the general population.

148 National and states HIV prevalence and seropositivity estimates were abstracted from national
 149 reports and all the available data sources respectively, between 2010 and 2014. All available PMTCT
 150 data (irrespective of contiguity of PMTCT facilities to ANC sentinel sites) were used to determine the
 151 state estimates. However, all four data sources of interest were not available for the same year due to
 152 the inconsistencies in the frequency at which national surveillance/surveys were conducted in Nigeria.
 153 The national surveys did not align with the years in which the RPD sources were available; thus, we
 154 paired the NSDs and the RPDs based on the closest years of availability (Table 1A).

Table 1A. Collated Data Sources for HIV Prevalence Estimates in Nigeria

Data Categories	Data Sources	Years Available				
		2010	2011	2012	2013	2014
NSD	ANC	Yes	N/A	No	N/A	Yes
	NARHS	N/A	N/A	Yes	N/A	N/A
RPD	PEPFAR APR HTC	N/A	N/A	N/A	Yes	Yes
	PEPFAR APR PMTCT	N/A	N/A	N/A	Yes	Yes

Note: No means that the data was expected for the year but not available or not in the format useful for the triangulation analyses. NA means the data was not expected for the year based on the original periodic design. None of the data sources were applicable (NA) in 2011, so the year was removed from the table.

155

156 Hence, the data sources were batched in the order of most matched or contiguous years and
 157 Batch 1 was most matched data sources (Table 1B).

158 Table 1B. Batch grouping of the HIV Data Sources by year of availability for analyses

Batch Number	Data Sources	
	Source Name	Year
1	ANC-HSS	2014

	PEPFAR-HTC	2014
	PEPFAR-PMTCT	2014
2	NARHS	2012
	PEPFAR-HTC	2013
	PEPFAR-PMTCT	2013
3	ANC-HSS	2010
	PEPFAR-HTC	2013
	PEPFAR-PMTCT	2013

Data Analysis

Seropositive estimates were used as determined by the PEPFAR/HIV program. There were no advanced or further statistical adjustments to the methodologies or analysis. The program simply divided the number of HIV positive clients by the total number tested by states and any other sub national units (SNUs). Prevalence estimates from NSD sources were collated as reported in the national reports. We do not intend to manipulate any process or methods that yielded the estimates.

Visual analyses were performed to determine the geographic distribution patterns and concurrence levels of HIV prevalence and seropositivity using maps, and the level of correlation between the paired sources using the scatter plots charts. The scatter plot graphs were used to visualize the extent of correlation between the absolute prevalence estimates of the RPD sources and the NSD sources. A linear regression model line was fitted to show relationship pattern as well as determine the correlation factor (R^2) of the paired data sources. We used ArcMap 10.2 to generate visual maps and geographic distribution concurrence levels, all set at quantile classifications with four classes. The distribution concurrence levels were computed as the proportion of states (between the two compared data sources) that showed exactly the same HIV prevalence color patches. Technically, this means the number of states (between the paired sources in comparison) that fall within the same HIV prevalence quartile divided by the 37 (the 36 states plus Federal Capital Territory). These analyses were performed to respond to the first objective of the study.

178 Bivariate simple linear regression analysis were performed to determine the measures and
179 strengths of the predictive relationships between the respective paired data sources in response to
180 the third study objective . The SPSS-IBM 21.0 was used to perform the correlations and line charts, and
181 linear regression modeling. For the linear regression, the RPD sources were used as the independent
182 variables (predictors) and NSD sources as the dependent (response) variables, and leaving the entry
183 method at “Enter”. This means that all the predictor/independent variables were entered into the
184 model at once. Significance was assessed at $P < 0.05$. Data sources from program and national surveys
185 were analyzed in pairs to facilitate clarity within batches.

186 Trend analysis was performed on the respective national estimates (median of state-level
187 estimates) from the data sources to determine the concurrence on their reflection of the trend of HIV
188 prevalence in the country within the study period.

189 Results

190 Analysis for Batch 1 data sources (ANC-HSS 2014, HTC 2014 and PMTCT 2014) showed that the
191 median prevalence was lowest in the PMTCT 2014 (1.3%) and highest in the HTC 2014 (3.5%). The PMTCT
192 data showed the most compact distribution (standard error[SE] 0.16), followed by HTC (SE 0.33) and
193 ANC-HSS (SE 0.5). Batch 2 (NARHS 2012, PMTCT 2013, HTC 2013) showed similar pattern for HTC and
194 PMTCT median prevalence, with HTC (4.2%) more than double that of PMTCT (1.6%). The NARHS 2012
195 survey had a median prevalence of 2.3%. Again, the PMTCT data showed a similar compact distribution
196 (Standard deviation [SD] 1.91) than the other data sources (NARHS: SD 3.34; HTC: SD 2.16) as with
197 Batch 1 data sources. In Batch 3, median prevalence of ANC-HSS 2010 was 4.1%. These estimates stand
198 between the lowest PMTCT 2013 and highest HTC 2013 median rates.

Table 2. Summary State Level HIV Prevalence Distribution from Routine and Surveillance Data Sources in Nigeria (2010-2014) n = 37

		ANC-HSS 2014	PMTCT 2014	HTC 2014	NARHS 2012	PMTCT 2013	HTC 2013	ANC-HSS 2010
Mean (SEM)		4.00 (0.50)	1.50 (0.16)	3.91 (0.33)	3.45 (0.55)	2.28 (0.31)	4.84 (0.35)	4.80 (0.46)
Median		3.0	1.28	3.47	2.30	1.62	4.20	4.10
Std. Dev.		3.03	0.96	2.01	3.34	1.91	2.16	2.80
Min		0.9	0.3	1.4	0.2	0.6	1.8	1.0
Max		15.4	4.7	10.4	15.2	9.2	11.0	12.7
Perc.	25	1.90	.76	2.65	1.05	1.15	3.35	2.50
	75	5.45	2.10	4.92	5.45	2.35	6.40	6.55

Concurrence and Correlation Patterns

Figures 1-3 show line charts comparing states' mean HIV prevalence rates between the data sources within each batch. Batch 1 graphs (Figure 1A&B) show that the ANC-HSS 2014 has a more correlated pattern with PMTCT 2014 at $R^2 = 44.0\%$ than with the HTC 2014 at $R^2 = 23.3\%$

Fig 1. Scatterplot of ANC-HSS 2014 and RPD sources 2014

Figure 2, compares the mean HIV prevalence rates for the NARHS 2012, a National population based survey with PMTCT 2013 and HTC 2013 seropositivity rates. The NARHS 2012 source demonstrated low correlation patterns with the respective RPD sources at R^2 less than 17%; however, it shows better correlation with HTC 2013 at $R^2 = 16.5\%$ than with PMTCT 2013 at $R^2 = 14.6\%$.

Fig 2. Scatterplot of NARHS 2012 and RPD sources 2013

Figure 3 shows that the correlation patterns of batch 3 paired data sources were similar to that of batch 2; however, with a stronger correlation factors between the paired ANC-HSS and RPD sources ($R^2 = 55.7\%$ between ANC-HSS 2010 and PMTCT 2013 and $R^2 = 35.9\%$ between ANC-HSS 2010 and HTC 2013).

Fig 3. Scatterplot of ANC-HSS 2010 and RPD sources 2013

Geographic Distribution Patterns of HIV Prevalence / Seropositivity rates

217 The geographic HIV prevalence patterns of the batch 1 data sources were similar (Fig 4). All show
218 common higher HIV prevalence in the lower South-West, South-South, South-East, eastern part of the
219 North-Central, and part of North-East regions of the country. ANC-HSS 2014 and PMTCT 2014 data from
220 batch 1 had 40.5% (n=15)¹ concurrence, and ANC-HSS 2014 and HTC 2014 data had 35.1% (n=13)²
221 concurrence. Generally, four states (Akwa-Ibom, Benue, Nassarawa, and Rivers States) had consistent
222 prevalence patterns across the three data sources

223 **Fig 4. Map Geographic Distribution Pattern of HIV Prevalence from Batch 1 Data Sources.**

224
225 Batch 2 data sources showed a similar regional pattern as in batch 1 (see Fig 5). The distribution
226 concurrence between NARHS 2012 and PMTCT 2013 was 40.5% (n = 15)³ and 37.8% (n=14)⁴ for NARHS
227 2012 and HTC 2013.

228 **Fig 5. Map Geographic Distribution Pattern of HIV Prevalence from Batch 2 Data Sources.**

229 Comparing ANC-HSS 2010 to PMTCT 2013 and HTC 2013, the concurrence between ANC-HSS 2010 and
230 PMTCT 2013 was 46% (n=17)⁵ and 37.8% (n=14)⁶ with HTC 2013 (Figure 6).

231 **Fig 6. Map Geographic Distribution Pattern of HIV Prevalence from Batch 3 Data Sources.**

232 **Analysis of Correlations and Associations between the Data Sources**

233 In batch one, there was a significant correlation between mean HIV prevalence from PMTCT2014
234 program data and ANC-HSS 2014 surveillance report ($R = 0.7, p < 0.01$); and between the HTC2014
235 program HIV prevalence data and ANC-HSS 2014 surveillance report ($R = 0.4, p < 0.05$) as shown in table

¹ ANC-HSS 2014:PMTCT 2014 states with similar pattern (Kebbi, Zamfara, Jigawa, Yobe, Oyo, Ekiti, Bayelsa, Rivers, Akwa Ibom, Ebonyi, Benue, Nasarawa, FCT, Enugu and Adamawa).

² ANC-HSS 2014:HTC 2014 states with similar pattern (Kebbi, Zamfara, Jigawa, Bauchi, Oyo, Edo, Delta, Bayelsa, Rivers, Akwa Ibom, Benue, Nasarawa and Ogun).

³ NARHS 2012:PMTCT 2013 states with similar prevalence patterns (FCT, Nassaraw, Akwa Ibom, Rivers, Taraba, Cross-River, Osun, Jigawa, Kwara, Enugu, Bauchi, Katsina, Zamfara, Kebbi, and Ebonyi).

⁴ NARHS 2012:HTC 2013 states with similar prevalence patterns (Katsina, Zamfara, Kebbi, Osun, Ekiti, Edo, Kogi, Enugu, Enonyi, Cross River, Rivers, Akwa Ibom, Taraba, and Plateau)

⁵ ANC-HSS 2010:PMTCT 2013 states with similar prevalence patterns (FCT, Nassaraw, Benue, Akwa Ibom, Rivers, Abia, Borno, Kogi, Lagos, Niger, Oyo, Ogun, Yobe, Bauchi, Katsina, Zamfara, and Kebbi)

⁶ ANC-HSS 2010:HTC 2013 states with similar prevalence patterns (Benue, Akwa Ibom, Abia, Bayelsa, Kaduna, Lagos, Niger, Ogun, Delta, Katsina, Zamfara, Kebbi, Kwara, and Ekiti)

236 3. The correlation between the HTC2014 program data and ANC-HSS 2014 data was statistically
237 significant; however, it was rather weaker than the PMTCT2014 - ANC-HSS2014 correlation. In batch
238 two, the individual correlations between the mean HIV prevalence of the two program data sources
239 (PMTCT 2013 and HTC 2013) and the population based surveillance report (NARHS2012) were positive
240 and statistically significant. PMTCT2013 and NARHS2012 was correlated as $R = 0.38$ ($p < 0.05$). HTC2013
241 and NARHS2012 was correlated at $R = 0.41$ ($p < 0.05$). The pattern of correlations among the paired
242 sources in batch 3 were quite similar to that of batch 1. However, batch 3 correlations showed overall
243 stronger statistically significant correlations between the 2013 RPD sources and ANC2010 report at the
244 individual paired levels. The correlation between PMTCT2013 and ANC2010 was stronger ($R = 0.75$, $p <$
245 0.001) compared to that between HTC2013 and ANC2010 ($R = 0.60$, $p < 0.001$).

246 **Overall Comparative National HIV Prevalence Trend from Routine and** 247 **Program Data Sources in Nigeria**

248 On further review of the data sources to understand the HIV prevalence trend in the last three
249 years among the three data sources with more than one time point for comparison (ANC-HSS, PMTCT
250 and HTC), all suggested a steady decline in the HIV prevalence from 2012 to 2014 for PMTCT and HTC,
251 and 2010 to 2014 for ANC-HSS (no middle year estimate was available) (Figure 7). However, the
252 prevalence estimates from the HTC program data were consistently higher, while the prevalence
253 estimates from the PMTCT program data were lower prevalence over the period. The ANC-HSS
254 estimates, though interpolated on two extreme estimates, takes the middle values with HTC estimates
255 providing the upper range and PMTCT providing the lower range along the trend.

256 **Fig 7. Trend Pattern of HIV Prevalence Estimates between 2010 and 2014 from the three data**
257 **sources.**

258

259

Predictive Association between Paired Data Sources

Table 3 also shows that the predictive association between PMTCT2014 program data and the ANC2014 HIV prevalence data in batch 1 was statistically significant at $\beta = 2.09$, (Confidence Interval [CI] 1.28, 2.90; $R^2=43\%$), and the association between the HTC 2014 and ANC-HSS 2014 prevalence estimates was also statistically significant at $\beta = 0.6$, (CI 0.10, 1.06; $R^2=12\%$). In batch 2, the association between NARHS2012 and the individual RPD sources were $\beta = 0.67$ (CI 0.11, 1.22; $R^2=12\%$) for PMTCT2013 and $\beta = 0.63$ (CI 0.14, 1.11; $R^2=14\%$) for HTC2013. In batch 3, the PMTCT2013 showed statistically significant association with ANC2010 at $\beta = 1.09$ (CI 0.76, 1.42; $R^2=57\%$) and HTC2013 showed similar association with ANC2010 but at $\beta = 0.78$ (CI 0.43, 1.13; $R^2=35\%$).

Table 3. Correlation and Regression Models of the paired RPD and PMTCT data sources

Batch #	Dependent variable	Model (Predictors)	Pearson Correlation	β beta coeff. (Std.Error)	R2	95% Confidence Interval for β	
						Lower Bound	Upper Bound
1	ANC-HSS 2014	PMTCT2014	0.66***	2.09*** (0.39)	43%	1.28	2.90
		HTC2014	0.38**	0.58* (0.24)	12%	0.10	1.06
2	NARHS 2012	PMTCT2013	0.38*	0.67* (0.27)	12%	0.11	1.22
		HTC2013	0.41**	0.63** (0.24)	14%	0.14	1.11
3	ANC-HSS 2010	PMTCT2013	0.75***	1.09*** (0.16)	57%	0.76	1.42
		HTC2013	0.60***	0.78*** (0.17)	35%	0.43	1.13

$n = 38$. $p \leq 0.001$ ***, $p \leq 0.01$ **, $p \leq 0.05$ *

Discussion

Our findings show that PMTCT data sources aligned more closely with the ANC-HSS data sources than HTC data sources did with the same ANC-HSS. This alignment pattern is understandable considering the fact that the ANC-HSS and PMTCT share similar population, which is pregnant women. On the other hand, though the PMTCT2013 and HTC2013 data demonstrated no appreciable correlation with the NARHS2012 source, and similar wideness to their fitted lines ($R\text{-squared} < 17\%$), their patterns showed

276 a positive relationships with the NARHS data. No clear differences in terms of individual visual alignments
277 could be observed between the two RPD sources and NARHS; however, the ANC-HSS data seem to show
278 a better visual alignment with PMTCT data sources than HTC sources .

279 The alignment pattern analysis of the state prevalence estimates in the different data batches
280 were further supported by the comparative geographic distribution map analysis of the paired data
281 sources. On comparing the geographic distribution pattern and concurrence levels of the RPD and NSD
282 sources among the sub-national units, the PMTCT RPD and the corresponding NSD sources
283 demonstrated higher concurrence levels across the three batches than seen with the HTC RPD sources.
284 For example, the PMTCT RPD source concurrence level from batch 1 to 3 were 40.5%, 40.5%, and 46%
285 compared to 35.1%, 37.8%, and 37.8% with HTC RPD sources. The map review, in addition, showed
286 similarities in the general regional distribution pattern in HIV prevalence estimates between the paired
287 data sources. They demonstrated strong evidence that the higher burden of the disease may reside at
288 the South-Eastern, Central-Eastern, and extreme North-Eastern Nigeria. This further demonstrated
289 that the PMTCT data could provide similar regional estimates of the disease burden than the HTC data
290 with reference to the ANC HIV sero-prevalence sentinel survey estimates. In all, the map analysis
291 strengthens the similarities and utility of RPD in monitoring the country prevalence using four
292 classification levels of the prevalence.

293 In all instances, PMTCT estimates had stronger significant association with both ANC-HSS and
294 NARHS estimates compared with HTC, but more with ANC-HSS than NARHS. This could be explained by
295 the common target population from which the ANC-HSS and PMTCT programmatic data sources were
296 generated, and similar in methods and settings. These findings were consistent with the findings
297 observed in similar studies in India, [19] Rwanda, [10] and Mozambique [11]. However, the utility of
298 PMTCT data in substituting ANC-HSS data had been questioned, largely due to perceived poor quality in

299 the testing and data management practices at PMTCT program sites [12-14]. Bolu et al., however,
300 suggested that despite these reservations, the PMTCT still has better utility in HIV program management
301 when used in conjunction with the ANC-HSS data [6].

302 Our findings demonstrated a decreasing trend in the national seropositivity estimates from 2012
303 to 2014. This trend is similar to that of the national HIV prevalence observed in ANC-HSS source from
304 2010 to 2014; however, only two estimates in 2010 and 2014 were available. The trend alignment is
305 explained by the higher to lower mean seropositivity rate / prevalence derived from these sources. The
306 trend from NARHS could not be ascertained because only one prevalence estimate in 2012 was available.
307 Furthermore, the relatively low prevalence estimates from the PMTCT data sources (compared to ANC-
308 HSS) in this study were similar to those reported in the literature. Earlier studies found that the PMTCT
309 prevalence data were general lower than that of the ANC-HSS [6,10,11] except Seguy et al. who found
310 the PMTCT prevalence (14%) a little higher than the ANC-HSS prevalence (13%) [9]. Our results suggested
311 that PMTCT prevalence estimates may be lower than the national survey prevalence while the HTC
312 prevalence estimates are higher than the national surveys; thus, PMTCT prevalence may form the lower
313 range of the national surveys while the HTC estimates forms the upper range limits.

314 Although our study data sources did not include sufficient population-based survey data, the
315 results concur with the spectrum suggested that the ANC-HSS may overestimate the prevalence in the
316 general population but could still reliably estimate the HIV prevalence trend in the general population.
317 The suggestion is consistent with the conclusions from similar comparative studies in 26 countries with
318 generalized epidemics, [12] in Zimbabwe, [13] in sub-Saharan Africa [14] and in Uganda [15] in which the
319 ANC-HSS estimated prevalence were slightly higher than the prevalence observed in the population-
320 based surveys. In their detailed review of the correlations, Gouws et al. suggested that the ANC-HSS
321 data be multiplied by 0.8 to adjust for epidemic trend in the general population [12]. There is an essential

322 difference between the NARHS and ANC-HSSs, either in their respective primary target populations or
323 the methodology or both.

324 The HTC data is generated from across the entire population groups by age and sex but only from
325 those who visited the health and mobile community testing facilities, unlike the NARHS data that is
326 generated across the entire population groups from the households and through scientific sampling. This
327 could explain the wide differences in the similarity of the program data with the NARHS, however, in all,
328 NARHS data showed a slightly closer data fitting with the HTC than the PMTCT data. We observed that
329 the HTC data overestimated the ANC-HSS prevalence, in contrast to the under-estimation of the PMTCT
330 data. The declining trend pattern observed across the ANC, HTC, PMTCT prevalence estimates further
331 shows that the HIV epidemic in Nigeria may be on the decrease. The three data sets all agree on this
332 pattern and further strengthen the potential use of the raw programmatic data for monitoring of the
333 country's HIV epidemic trend.

334 The observed correlations agree with our modelled results to a large extent. Our finding
335 suggests that PMTCT program data could reflect or closely monitor state level epidemic pattern
336 obtainable with the ANC-HSS data, but may not produce exact point estimates. Our findings suggest that
337 the PMTCT seropositivity may be multiplied by 2.0 to adjust for ANC-HSS prevalence estimates.
338 Generally, the findings agree with that of most previous related studies, pulling more weight to the
339 usefulness of PMTCT data. This further reveals the potential direction of research toward the exploration
340 of the use of other RPD from other indicators and disease programs for local health surveillance. Such
341 future studies could be a break-through for less-expensive, easy, and fairly reliable m for health
342 surveillance in resource-limited countries.

343 **Limitations**

344 This study is subject to at least four limitations. First, except for batch 1 data sources, other
345 batch data sources were not from the same year; thus, their prevalence estimates may not be a perfect
346 match. Secondly, the RPD were generated from the health facilities across the country; however, the
347 facility distribution pattern of the country was not determined in this study, hence, could not
348 determine the representativeness of the RPD to inform the appropriateness of estimating population
349 based surveys such as NARHS and DHS [7,20]. The implication of this is that the potential utility of the
350 RPD for monitoring national HIV epidemic pattern may change with change in the distribution of the
351 facilities that generate these data. These findings assume that facilities were fairly distributed all over
352 the country. In addition, the quality of the respective methodologies, HIV testing services in the
353 facilities, and data managements of the data sources could not be determined in this study and may
354 improve the correlation analyses of the data sources. Lastly, the concurrence analysis that quantified
355 the geographic distribution accounted for the exact concurrence. This means that a state with level 4
356 and 3 prevalence between two data sources were considered as non-concurrence in the analysis
357 regardless of the closeness. Such closeness was considered same as level 4 and 1 between the data
358 sources. However, direct visual review of the maps could help resolve this limitation.

359 **Conclusions**

360 The study findings suggest that the RPD sources could be significantly correlated and associated
361 with the NSD sources in monitoring the HIV prevalence in the country. The RPD sources could reliably
362 demonstrate the trend in HIV epidemic at national and state levels. As expected, PMTCT data tend to
363 underestimate the ANC-HSS HIV prevalence while the HTC data showed indication of overestimation of
364 the epidemic trend among the reproductive female population. The potential use of routine PMTCT
365 programmatic data may be more reliable and favorable than HTC program data in estimating the
366 national HIV epidemic among women of reproductive age in Nigeria. The implication of the finding is

367 that, PMTCT could largely be a more promising source for HIV surveillance in the local settings in Nigeria
368 and other resource-limited countries, particularly in settings where no monitoring information exists or
369 periodic population-based surveys are infrequent.

370 **Acknowledgements**

371 This research [or project, publication, etc.] has been supported by the President's Emergency Plan for
372 AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC)

373 The authors acknowledge and appreciate the thorough reviews and edits made by Dr. Stacie Greby.

References

1. The Joint United Nations Programme on HIV and AIDS. Nigeria HIV and AIDS Estimate 2016 [Internet]. 2019 [cited 20 Feb 2019]. Available: <http://www.unaids.org/en/regionscountries/countries/nigeria>
2. The Joint United Nations Programme on HIV and AIDS. HIV and AIDS Estimate 2016 - Nigeria Country Factsheet [Internet]. 2019 [cited 20 Feb 2019]. Available: <http://www.unaids.org/en/regionscountries/countries/nigeria>
3. Awofala AA, Ogundele OE. HIV epidemiology in Nigeria. *Saudi J Biol Sci.* 2018;25: 697–703. doi:10.1016/j.sjbs.2016.03.006
4. Nigeria Federal Federal Ministry of Health (FMOH) Nigeriaof Health, Nigeria. National HIV Sero-prevalence Sentinel Survey among Pregnant Women Attending Antenatal Clinics 2010 [Internet]. 2011 [cited 12 Nov 2016]. Available: http://nepwhan.net/assets/ANC_2010_Report.pdf
5. Federal Ministry of Health (FMOH) Nigeria. National HIV Sero-prevalence Sentinel Survey among Pregnant Women Attending Antenatal Clinics in Nigeria 2014 [Internet]. 2015 [cited 12 Nov 2016]. Available: http://nigeriahealthwatch.com/wp-content/uploads/bsk-pdf-manager/1176_2014__National_HIV_Sero-prevalence_Sentinel_Survey_among_Pregnant_Women_Attending_Antenatal_Clinics_in_Nigeria._FMOH_1229.pdf
6. Bolu O, Anand A, Swartzendruber A, Hladik W, Marum LH, Sheikh AA, et al. Utility of antenatal HIV surveillance data to evaluate prevention of mother-to-child HIV transmission programs in resource-limited settings. *Am J Obstet Gynecol.* 2007;197: S17-25. doi:10.1016/j.ajog.2007.03.082
7. Federal Ministry of Health (FMOH) Nigeria. National HIV & AIDS and Reproductive Health Survey 2012 [Internet]. 2013 [cited 12 Nov 2016]. Available: [http://nigeriahealthwatch.com/wp-content/uploads/bsk-pdf-manager/431_2012_National_HIV_&_AIDS_and_Reproductive_Health_Survey_\(NARHS_Plus_II,_2012\),_FMOH_Abuja_1172.pdf](http://nigeriahealthwatch.com/wp-content/uploads/bsk-pdf-manager/431_2012_National_HIV_&_AIDS_and_Reproductive_Health_Survey_(NARHS_Plus_II,_2012),_FMOH_Abuja_1172.pdf)
8. Federal Ministry of Health, Nigeria. Nigeria National PMTCT Guidelines 2010 [Internet]. 2010 [cited 12 Nov 2016]. Available: http://sbccvch.naca.gov.ng/sites/default/files/Nigeria_National-PMTCT-Guidelines_2010.pdf
9. Seguy N, Hladik W, Munyisia E, Bolu O, Marum LH, Diaz T. Can Data from Programs for the Prevention of Mother-to-Child Transmission of HIV be Used for HIV Surveillance in Kenya? *Public Health Rep.* 2006;121: 695–702.
10. Françoise Kayibanda J, Alary M, Bitera R, Mutagoma M, Kabeja A, Hinda R, et al. Use of routine data collected by the prevention of mother-to-child transmission program for HIV surveillance among pregnant women in Rwanda: opportunities and limitations. *AIDS Care.* 2011;23: 1570–1577. doi:10.1080/09540121.2011.579941
11. Young PW, Mahomed M, Horth RZ, Shiraishi RW, Jani IV. Routine data from prevention of mother-to-child transmission (PMTCT) HIV testing not yet ready for HIV surveillance in Mozambique: a retrospective analysis of matched test results. *BMC Infect Dis.* 2013;13: 96. doi:10.1186/1471-2334-13-96

- 411 12. Gouws E, Mishra V, Fowler TB. Comparison of adult HIV prevalence from national population-based
412 surveys and antenatal clinic surveillance in countries with generalised epidemics: implications for
413 calibrating surveillance data. *Sex Transm Infect.* 2008;84 Suppl 1: i17–i23. doi:10.1136/sti.2008.030452
- 414 13. Gregson S, Dharmayat K, Pereboom M, Takaruzza A, Mugurungi O, Schur N, et al. Do HIV prevalence trends
415 in antenatal clinic surveillance represent trends in the general population in the antiretroviral therapy era?
416 The case of Manicaland, East Zimbabwe. *AIDS Lond Engl.* 2015;29: 1845–1853.
417 doi:10.1097/QAD.0000000000000754
- 418 14. Montana LS, Mishra V, Hong R. Comparison of HIV prevalence estimates from antenatal care surveillance
419 and population-based surveys in sub-Saharan Africa. *Sex Transm Infect.* 2008;84 Suppl 1: i78–i84.
420 doi:10.1136/sti.2008.030106
- 421 15. Musinguzi J, Kirungi W, Opio A, Montana L, Mishra V, Madraa E, et al. Comparison of HIV prevalence
422 estimates from sentinel surveillance and a national population-based survey in Uganda, 2004-2005. *J*
423 *Acquir Immune Defic Syndr* 1999. 2009;51: 78–84. doi:10.1097/QAI.0b013e3181990713
- 424 16. Joint United Nations Programme on HIV/AIDS. National HIV estimates file [Internet]. 2018 [cited 3 Dec
425 2018]. Available: <http://www.unaids.org/en/dataanalysis/datatools/spectrum-epp>
- 426 17. Harklerode R, Schwarcz S, Hargreaves J, Boule A, Todd J, Xueref S, et al. Feasibility of Establishing HIV
427 Case-Based Surveillance to Measure Progress Along the Health Sector Cascade: Situational Assessments in
428 Tanzania, South Africa, and Kenya. *JMIR Public Health Surveill.* 2017;3. doi:10.2196/publichealth.7610
- 429 18. World Health Organization. HIV Triangulation Resource Guide - Synthesis of Results from Multiple Data
430 Sources for Evaluation and Decision-making [Internet]. Feb 2009 [cited 6 Oct 2018]. Available:
431 http://apps.who.int/iris/bitstream/handle/10665/44107/9789241597999_eng.pdf?sequence=1
- 432 19. Kumar R, Viridi NK, Lakshmi PVM, Garg R, Bhattacharya M, Khera A. Utility of Prevention of Parent-to-Child
433 Transmission (PPTCT) Programme data for HIV surveillance in general population. *Indian J Med Res.*
434 2010;132: 256–259.
- 435 20. National Population Commission. Nigeria Demographic and Health Survey 2013 [Internet]. 2014 [cited 12
436 Nov 2016]. Available: <https://dhsprogram.com/pubs/pdf/FR293/FR293.pdf>

437	Supporting information
438	S1 Fig. Scatterplot of ANC-HSS 2014 and RPD sources 2014
439	S2 Fig. Scatterplot of NARHS 2012 and RPD sources 2013
440	S3 Fig. Scatterplot of ANC-HSS 2010 and RPD sources 2013
441	S4 Fig. Map Geographic Distribution Pattern of HIV Prevalence from Batch 1 Data Sources
442	S5 Fig. Map Geographic Distribution Pattern of HIV Prevalence from Batch 2 Data Sources
443	S6 Fig. Map Geographic Distribution Pattern of HIV Prevalence from Batch 3 Data Sources
444	S7 Fig. Trend Pattern of HIV Prevalence Estimates between 2010 and 2014 from the three data
445	sources.
446	S1 Table. Study dataset