



Published in final edited form as:

J Acquir Immune Defic Syndr. 2021 August 01; 87(Suppl 1): S43–S51. doi:10.1097/QAI.0000000000002704.

Improving Sampling Efficiency for Determining Pediatric HIV Prevalence in National Surveys: Evidence from Eight Sub-Saharan African Countries

Giles Reid, PhD^{1,*}, Andrew C. Voetsch, PhD², Paul Stupp, PhD², Stephen McCracken, PhD², Graham Kalton, PhD³, Sindisiwe Dlamini, MPH⁴, James McOlogi Juma, MSC⁵, Thoko Kalua, MBBS, MSc⁶, Wilford Kirungi, MBChB⁷, Masebeo Koto, MPH⁸, Owen Mugurungi, MD⁹, Lloyd Mulenga, MD¹⁰, Nicholus Mutenda, MD¹¹, Lawrence Marum, MD, MPH², Suzue Saito, PhD²

¹ICAP at Columbia University, New York, NY, USA

²U.S. Centers for Disease Control and Prevention, Atlanta, GA, USA

³Westat, Rockville, MD, USA

⁴Ministry of Health, Eswatini

⁵Ministry of Health, Community Development, Gender, Elderly and Children, Tanzania

⁶Ministry of Health and Population, Malawi

⁷Ministry of Health, Republic of Uganda

⁸Ministry of Health, Lesotho

⁹Ministry of Health and Child Care, Zimbabwe

¹⁰Ministry of Health, Zambia

¹¹Ministry of Health and Social Services, Namibia

Abstract

Background—Measurement of mother to child HIV transmission through population-based surveys requires large sample sizes because of low HIV prevalence among children. We estimate potential improvements in sampling efficiency resulting from a targeted sample design.

Setting—Eight countries in sub-Saharan Africa with completed Population-based HIV Impact Assessment (PHIA) surveys as of 2017.

Methods—The PHIA surveys used a geographically stratified two-stage sample design with households sampled from randomly selected census enumeration areas. Children (0–14 years of age) were eligible for HIV testing within a random subsample of households (usually 50%).

*Corresponding author: Address: ICAP at Columbia University, Mailman School of Public Health, 722 West 168th Street, New York, NY 10032, Phone: +1 (917) 385 7099, Fax: +1 (212) 342 1824, giles.reid@columbia.edu.

Disclaimer

The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the funding agencies.

Estimates of child HIV prevalence in each country were calculated using jackknife replicate weights. We compared sample sizes and precision achieved using this design to a two-phase disproportionate sample design applied to strata defined by maternal HIV status and mortality.

Results—HIV prevalence among children ranged from 0.4% (95% CI: 0.2–0.6) in Tanzania to 2.8% (95% CI: 2.2–3.4) in Eswatini with achieved relative standard errors (RSEs) between 11–21%. The expected precision improved in the targeted design in all countries included in the analysis, with proportionate reductions in mean squared error (MSE) ranging from 27% in Eswatini to 61% in Tanzania, assuming an equal sample size.

Conclusions—Population-based surveys of adult HIV prevalence that also measure child HIV prevalence should consider targeted sampling of children to reduce required sample size, increase precision, and increase the number of positive children tested. The findings from the PHIA surveys can be used as baseline data for informing future sample designs.

Keywords

children; adolescents; HIV/AIDS; sample design; survey; PHIA

Introduction

The elimination of new HIV infections among children is a major focus of the global HIV response^{1–3}. Better measurements of pediatric prevalence and the burden of HIV infection allow for better allocation of resources, tracking of treatment impact, and mortality reduction^{4,5}. The Population HIV Impact Assessment (PHIA) household surveys carried out since 2015 have provided a wealth of information, including HIV prevalence and treatment among children. However, the sampling method used, where children were tested for HIV in a random subsample of the surveyed households, requires a large subsample size, because HIV is relatively rare among children even in countries with high adult prevalence. This has contributed to a longstanding dearth of population-level pediatric HIV data. Program planning typically relies on estimates from models based on births to HIV-infected pregnant women, treatments they and their infants receive, and survival estimates of these treated and untreated children using assumptions from national surveys and program data. These models have many potential sources of error: prevalence, fertility and other characteristics of the pregnant women; adherence and effectiveness of the treatments given; incidence of HIV during pregnancy and lactation; and survival curves of infected children who receive or do not receive treatment. Direct estimates of pediatric prevalence enable validation and correction of model assumptions and parameters⁶.

Several survey sampling methods have been developed for estimation of rare traits using screening to identify sub-populations with higher rates of these traits and targeting the sample towards these groups^{7,8}. In many survey settings, the time, cost, and practical difficulties involved in the initial screening outweigh theoretical improvements in precision or sample size. It is also common for the gain in precision from these approaches to be modest because the population is not concentrated enough in identifiable sub-populations that can serve as the strata⁹. However, in PHIA surveys, maternal HIV status and maternal mortality is immediately available at no additional cost from the adult data collection.

Since pediatric HIV is predominantly vertically transmitted¹⁰, this suggests that sampling efficiency in national surveys that also measure HIV in adults could be increased by subsampling of children who are more likely to have HIV based on maternal HIV status.

In this paper, we use the results of recent PHIA surveys in eight countries in sub-Saharan Africa to determine whether using the HIV status of mothers to select higher-risk children could reduce the sample size needed to achieve a certain level of precision, compared with the household-level random subsampling design used in PHIA and other recent general-population HIV surveys.

Methods

Study design

PHIA surveys are designed to provide national estimates of HIV incidence and subnational estimates of HIV prevalence and viral load suppression to assess the HIV epidemic and the impact of HIV prevention and ART programs¹¹. As well as PHIA, several other surveys have been carried out using similar household sampling designs^{12–14}. These surveys have also enabled measurement of pediatric HIV prevalence and progress towards the goal of eliminating new HIV infections among children¹⁵. Sample designs were stratified in each country by major geographic subdivisions used in the national health system or census. Census Enumeration Areas (EAs) within each stratum were selected with probability proportional to their number of households in the most recent census. In each selected EA, households were enumerated by field staff, and the final systematic sample of households was drawn to give an equal overall probability sample of households within each stratum. All adults above age 15 in Tanzania and Eswatini, between 15 and 59 years of age in Lesotho and Zambia, and between 15 and 64 years in all other countries in the sampled households were selected. All children (0–14 years of age) were selected in a subsample of these households. This paper uses results from the PHIA surveys conducted in Zimbabwe, Zambia, Malawi, Eswatini, Uganda, Namibia, Lesotho, and Tanzania between 2015 and 2017. See the PHIA project documentation for more detail about the PHIA surveys included in this study^{16–18}.

In each country, the PHIA survey design included the selection of a random sub-sample of households for HIV testing of children. Only children in these sub-sampled households were eligible for blood draw and HIV testing. In Tanzania 33% of households were selected for testing children. In Uganda 60% of households were selected for testing of children 0–4 years, with 34% of those households also selected for testing of children 5–14 years. In all other countries a 50% household sub-sample was taken. The sampling percentages were based on the sample size required to meet all of the country-specific sample design targets.

Study population

All resident adults over the age of 15 years, but younger than a country-specific maximum age, who slept in a selected household the night before the interview were eligible for both interview and HIV testing.

Children in the sub-sampled households who had slept in the household the night before the survey were eligible for HIV testing. Consent from the child's parent or guardian was required before the child was tested, and for older children above a country-specific age (ranging from 9 to 12) assent was also needed from the children themselves.

Data collection

Household and individual questionnaires were administered using computer-assisted personal interviewing. The adult interview included demographics, HIV testing and treatment history, and questions about the respondents' children in the household. HIV testing and counseling was done during the survey in private locations in or near the participants' residences, using each country's national HIV rapid testing algorithm. The test results were immediately returned to participants. We use these rapid tests results as the screening data in the analyses reported below. Further blood tests, including HIV viral load and confirmatory PCR HIV tests for infants under 18 months of age, were carried out in a laboratory using plasma samples or dried blood spots.

Analysis

Survey weights—Weights for all PHIA surveys were calculated based on the initial selection probabilities, adjusted for nonresponse and under coverage. Nonresponse adjustments used “weighting adjustment cells” defined by survey variables chosen to predict response and HIV status. These variables were selected using a two-step method: first, a Least Absolute Shrinkage and Selection Operator was used to create an initial set of variables most predictive of survey response, and then chi-square automatic interaction detection was used to build and apply a nonresponse adjustment model using these variables.

To correct for under coverage, post-stratification adjustments were made so that the sum of the weights in each five-year age group by gender cell matched the most recent available national population projections. The resulting weights were used to estimate child HIV prevalence. Variances, design effects, and confidence intervals were estimated using jackknife replicate weights in the SAS SURVEYFREQ procedure. These estimates are used as the basis for this paper's extrapolations to alternative sample designs in each country.

Mother-to-child linking—During the household interview, the household head identified all relationships between a child (0–14 years of age) and their mother or female guardian. During the adult interview, mothers are asked to confirm the relationship, so for each child there is either a single mother identified or a response from the household head that the mother is not living in the household. Implausible data, such as mothers under the age of ten, were removed during data cleaning¹⁶.

Estimation of sample size and precision under alternative sample designs—To compare the performance of alternative sample designs to the PHIA design, we used the data collected from PHIA households about the number of children present in each household and the mother's HIV status for each rostered child.

Sampling fractions for Neyman allocation—Using the adult survey data, eligible children in households sampled for the PHIA survey were allocated to one of three strata based on maternal HIV status: 1) mother tested HIV positive or not alive; 2) mother's HIV status unknown; and, 3) mother tested HIV negative. (for brevity, in the following we will refer to these strata as “positive/deceased”, “unknown”, and “negative”, respectively) The sampling fraction for the positive/deceased stratum was set to 1.0, and the proportions in the other strata were determined using Neyman allocation^{7,19}. For example, for the negative stratum, the sampling fraction is

$$v_{\text{negative}} = \frac{\sqrt{p_{\text{negative}}(1 - p_{\text{negative}})}}{\sqrt{p_{\text{positive}}(1 - p_{\text{positive}})}}$$

where the p 's are the prevalence rates in the negative and positive/deceased strata. The sampling fraction for the unknown stratum is calculated similarly. We adjusted the sampling fractions using the response rates within each maternal HIV stratum with the aim of achieving final respondent numbers which have the desired optimal ratios. The expected sample sizes in each stratum were calculated based on these sample fractions, the PHIA response rates, and household roster information.

Alternative designs

In addition to the Neyman allocation design (scenario 1), we analyzed two alternative designs. In one of these (scenario 2), we started from the Neyman allocation sampling fractions, set the sampling fraction for the negative stratum to zero, and adjusted the sample size in the unknown stratum to obtain the same total sample size. The second alternative (scenario 3) uses the PHIA design but with the household subsampling fraction set such that the sample size would be the same as scenarios 1 and 2. The equal sample sizes in the three scenarios allow the performance of the three alternatives to be directly compared.

We also computed the minimum sampling fraction required for the unknown stratum required to meet specific relative standard error (RSE) targets for each country, assuming that all children in the positive/deceased stratum were eligible. These results indicate the potential sample size reduction that could be achieved in the alternative design for a fixed precision.

Variance estimation

To estimate the variance of prevalence estimates in each alternative design we used PHIA estimates for prevalence within each maternal HIV status stratum and sample sizes for each design calculated in the previous step. First, the variance of the prevalence estimate was computed assuming simple random sampling within each stratum, using the following approximation for a two-phase disproportionate stratified design:

$$v(\bar{p}) \approx \sum \frac{w_h^2 p_h (1 - p_h)}{n_h} + \frac{1}{n'} \sum w_h (p_h - \bar{p})^2$$

where w_h , p_h , and n_h are the sample proportion, estimated prevalence rate, and sample size in stratum h respectively, n is the total sample size, and \bar{p} is the estimated total population child prevalence rate. The first sum relates to the variances in each stratum, and the second sum accounts for the variability in the stratum proportions caused by the household sampling.

To account for the effects of geographic clustering of children and weighting adjustments, we multiplied this variance by the achieved design effect for HIV prevalence among 0–14 year olds in the PHIA survey. This is an approximation but is adequate for our needs. For example, if the alternative design is much larger than the original child sample, the design effect may also be expected to be somewhat larger due to the greater clustering design effect arising from the increased average subsample size per primary sampling unit. However, the clustering design effect is limited because the intraclass correlation from HIV clustering by PSU is small.

Bias estimation and root mean squared error

In the design with no sample from the negative stratum, the assumption that the HIV prevalence among those children is zero introduces a bias. We estimated this bias using the overall weighted prevalence for this group from the combined data from all eight countries (0.05%) multiplied by the fraction of children with negative mothers.

This design has a lower total variance because the variance contribution from this stratum becomes zero, but the bias must be considered in comparing the overall accuracy to the other designs. We use the Root Mean Squared Error (RMSE) of the prevalence,

$$\text{RMSE} = \sqrt{\text{Bias}^2 + \text{Variance}}$$

and the relative RMSE (RRMSE), which is the RMSE divided by the estimated prevalence. For the Neyman allocation and PHIA designs we assume the bias is zero.

Results

HIV prevalence and achieved precision

For the PHIA subsampled household designs in each country, Table 1 summarizes the number of eligible children 0–14 years of age identified in responding households, the number who consented to blood draws and home-based HIV testing and counseling, overall response rates, and HIV prevalence rates by maternal HIV status. Overall, in the 49,649 households selected for child testing across the eight countries, we identified 69,237 eligible children, of whom 55,273 underwent blood draw and home-based HIV testing and counseling after a parent or guardian provided their consent. The overall unweighted response rate was 80%, ranging from 62% in Malawi to 96% in Uganda. Combining the data from all eight countries, there were 665 HIV positive children and the overall weighted HIV prevalence was 0.8% (95% CI: 0.7–0.9). HIV prevalence ranged from 0.4% in Tanzania to 2.8% in Eswatini. The RSEs varied by country from 11% in Eswatini and Zambia to just over 20% in Uganda and Tanzania.

Maternal HIV status

Overall, 57% of eligible children had an HIV-negative mother, with a range from 35% in Lesotho to 74% in Uganda, and 13% had either an HIV positive or deceased mother, ranging from 6% to 27% in the same countries. Maternal HIV status for the remaining 31% of eligible children was unknown (Table 1). The proportion of children in the unknown stratum varied widely by country, from 20% in Uganda to 46% in Namibia. The unknown stratum had the lowest response rate at 66%, while the positive/deceased and negative strata had similar response rates at 84% and 86%, respectively. This pattern was consistent across countries, except in Uganda where the response rate was over 95% in each stratum, and in Malawi where the positive/deceased stratum had a higher response rate than the negative stratum (78% vs. 72%).

There were three main reasons for unknown maternal HIV status: 1) non-response or refusal by eligible mothers; 2) mothers absent from the household at the time of the interview; and, 3) mothers reported as deceased by another member of the household. Less than two percent of unknown maternal HIV status was due to other reasons, such as inconclusive blood test results or data entry errors (Table 2). The proportion of unknown HIV status resulting from the three main reasons varied by country, especially for non-response/refusal by eligible mothers, which accounted for 11% of unknown status in Uganda, compared to 48% in Malawi. Mothers absent from the household at the time of the interview were responsible for 78% of the unknown maternal HIV status in Uganda and 44% in Malawi.

Child HIV prevalence by maternal status

Among children with unknown maternal HIV status, HIV prevalence was highest among children with a deceased mother (5.2%), followed by children of mothers who were not present during the interview (0.8%), and those whose mother had refused to participate (0.6%) (Table 2).

Overall HIV prevalence among children with either an HIV positive or deceased mother was 6.5% (95% CI: 5.6 – 7.5), with 484 children testing positive. By country, this figure ranged from 5.2% in Namibia to 8.5% in Malawi.

HIV prevalence among children of mothers who were alive but with unknown HIV status was 0.77% (95% CI: 0.54 – 0.99%), with a minimum of 0.15% in Tanzania and a maximum of 2.2% in Eswatini. There were 162 HIV positive children in this group.

A total of 19 HIV positive children had an HIV negative mother. Of these, 13 were female and 8 were aged under 10. The responding parent or guardian for six of these children knew their child's HIV positive status prior to the survey. We were not able to determine whether any of these children had incorrectly linked mothers or were known to have contracted HIV from someone other than their mother. Although 57% of all children tested had negative mothers, this stratum contained only 3% of the positive children. The overall HIV prevalence in this stratum was 0.05% (95% CI: 0.01 – 0.08%).

Precision estimates for targeted sample design scenarios

Table 3 shows the expected number of children tested and the precision (RSE) of child HIV prevalence estimates in each of the three sampling designs we considered. Under Neyman allocation with the sampling fraction in the positive/deceased stratum set to 1.0, around 10% of children with negative mothers and 16% (Tanzania) to 52% (Lesotho) of children with unknown status mothers would have been selected from the overall number of children rostered. The total number of children tested across these 8 countries, assuming PHIA response rates, would be approximately 42,000, 13,000 fewer than the actual PHIA surveys. Only the countries with the highest proportion of HIV positive mothers, Lesotho and Eswatini, would have had an increased sample size under the Neyman allocation design.

The scenario 1 design has a smaller expected RSE than the PHIA design where all children in a subsample (generally 50%) of households were selected. In most countries the sample sizes for the PHIA design and this design are similar. Exceptions occur with Tanzania, Uganda, and Zambia, where the PHIA sample sizes are much larger than those for the targeted design. These countries have the highest proportion of children with negative mothers, and had relatively low prevalence among the unknown stratum.

Under this proposed design, we would expect about 1,200 children to test positive, of whom 980 would be children with positive mothers. In each country, the expected number of positive children with negative mothers under this design would be between 0.1 and 1.7. In the PHIA surveys, a total of 665 children tested positive for HIV.

The scenario 3 design, which is the same as the PHIA design but with a sample size set equal to that of the scenario 1 and 2 designs, enables a direct comparison of the achieved precision assuming the same total number of children tested. Table 3 shows that the relative reduction in mean squared error using the targeted design ranges from 27% in Eswatini to 61% in Tanzania.

Estimated bias from sampling assuming zero prevalence among HIV negative mothers

The average sample size for the negative stratum is about 1,000 children per country, with an expected yield of around 1 positive child. Assuming a prevalence of zero and setting the sample size to zero allows an increase in the sample size in the unknown stratum but introduces an estimated bias in the overall prevalence in the range of 0.01% to 0.06% (Table 3, scenario 2). For a constant sample size, the overall RRMSE under this design is only marginally lower than the RSE in scenario 1. This is because the sample size in the negative stratum is relatively small, and the released sample can only be allocated to the unknown stratum, because all children in the positive/deceased stratum are already selected. In a more general design where the adult sample size is larger and the positive/deceased stratum sampling fraction is less than 1, the gains from this approach would be greater.

Required sample sizes to meet specified RSE targets

Primary objectives in PHIA were based on precision (RSE) targets. To adjust for this type of constraint, we can adjust the sampling fraction for the unknown stratum. The required sampling fractions and corresponding numbers of children tested are shown in Table 4 for a

range of target RSE values by country. To achieve a 20% RSE target a sampling fraction of less than 0.1 is needed in all countries except Uganda. The total number of children tested can be compared to the achieved RSEs and total sample sizes in Table 3. For example, under the PHIA design 10,345 children were tested in Uganda to achieve an RSE of 20%. The same RSE could have been obtained with a total sample size of less than 3,500 by testing all children in the positive/deceased stratum and 15% of those in the unknown stratum.

Discussion

Our results show that the precision of child HIV prevalence estimates could be improved by using a targeted sample design with a similar sample size as the PHIA design, and that for a given RSE target, substantial reductions in sample size are possible. For a given sample size, our design typically results in about a one-third reduction in mean squared error. This increase means that the same precision targets could be met for a substantially reduced cost, in comparison with the design used for the PHIA surveys. Because HIV in children is almost always acquired from the mother, targeting of the sample using maternal HIV status is effective across the wide range of epidemic contexts and prevalence parameters encountered in the eight countries analyzed.

Our design includes testing all children with positive mothers. In addition to the overall gains in efficiency, this design would yield almost twice the number of HIV positive children as the PHIA design with a similar number of children tested in total, providing a more detailed understanding of the population of HIV positive children. For example, as PMTCT programs have achieved greater effectiveness in reducing pediatric HIV infections, a higher proportion of infected infants are the result of undetected incidence of HIV in pregnant and breastfeeding women, and our targeted design could help understand the changes in transmission dynamics that have occurred^{20,21}. Among the children with positive mothers also will be many children whose HIV infection was previously unknown; linkage to treatment may result in a lifesaving benefit of survey participation.

The identification of 19 positive children with negative mothers is notable, though our data does not allow for any conclusions to be drawn about how these children were infected. It is likely that the parents or guardians of the six children whose HIV-positive status was known prior to the survey understand how the child became infected, but we did not ask for details about this in our questionnaire. In several cases there is some evidence that the true birth mother may have been either misreported or recorded incorrectly during the survey. During data cleaning we did correct cases in which there was clear evidence that the wrong household member was linked to the child, but in these remaining cases there is not enough information for us to be able to determine how the children were infected.

Regardless of the explanation for how children with reported HIV negative mothers contracted HIV, our results show that testing children with HIV negative mothers provides little benefit in the estimation of overall pediatric prevalence. Our results also represent a useful upper limit on the true non-vertically-transmitted HIV prevalence among children in the surveyed countries. Given the potential for mis-reporting of a child's birth mother and the extremely small number of positive children found in this group, we consider it

likely that our results are an overestimate of the true bias. Assuming zero prevalence among these children and re-allocating resources to the unknown stratum provides a slightly smaller overall RRMSE, with only a small bias, while simplifying the design.

We also analyzed the reasons for unknown maternal statuses. Although the overall percentage of children with unknown maternal status is similar in most of the studied countries, the explanations show a wide variation in the distribution between non-response/refusal and mothers not being present in the household for the survey. Reducing the rate of unknown maternal statuses would result in further increases in sampling efficiency, so further studies of this kind should consider how survey procedures can encourage the participation of mothers, for example ensuring that convenient household revisit times are available and providing specific information about the benefits of enrolment to reduce non-response and refusal. Our findings could assist in determining which measures will be most important in the context of individual countries.

This study has several limitations. Most importantly, the sample designs are built upon a household survey where all mothers are eligible for testing, and where initial maternal results are available during fieldwork, so that maternal HIV status can be used as a screening variable for no additional cost.

This study also relies on extrapolating results from one sample design to an alternative design. Some parameters, such as response rates, could differ under the alternative design. The impact on study participants themselves of changing the design is likely minimal, since there would be no obvious change in their experience of the survey, but we have very limited evidence about these practical aspects.

There are challenges to implementing a targeted pediatric HIV testing approach. If all children in the positive/deceased stratum and no children in the negative stratum are sampled, only the sampling fraction for the unknown stratum needs to be determined. Allowance needs to be made for the lower response rate in this stratum observed in the PHIA surveys. Table 4 shows that similar RSEs to those achieved using the PHIA surveys' subsampling design could have been obtained using a sampling fraction in the unknown stratum of less than 50%. For a country lacking country level specific data to justify an alternative choice, a 50% sampling fraction is likely to lead to a smaller sample size than the PHIA design while reaching a comparable RSE and will be robust to a wide variation in prevalence and stratum proportions.

If the main survey sample size is very large, the number of children selected may be larger than desirable since all children are potentially eligible. To control the sample size, the design can be modified to use a sampling fraction of f in the positive/deceased mother stratum and $f/2$ in the unknown stratum where a mother is absent or refuses testing. In general, controlling sample size requires estimates of the proportions of children in each of the three strata. The findings on these proportions reported in Tables 3 and the corresponding sampling fractions in Table 4 may provide useful guidance.

Future population-based surveys of adult HIV that also plan to measure child HIV prevalence should consider targeted sampling to reduce the sample size needed to meet

a precision target. The targeted design also provides a much higher yield of both positive children and children of HIV positive mothers for a fixed sample size, allowing for more granular analysis of pediatric HIV prevalence and prevention of mother to child HIV transmission as well as potential lifesaving linkage to care and treatment for infected children. The findings from the PHIA surveys can be used as baseline data for future sample designs.

Source of support:

This research has been supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the U.S. Centers for Diseases Control and Prevention under the terms of Cooperative Agreements 1U2GGH000994 and 5NU2GGH001226.

References

1. Blais P, Sirivar S, Seto J. Supporting implementation research to improve coverage and uptake of HIV related interventions. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2017;75:S109–S110. [PubMed: 28498178]
2. UNAIDS. Start Free, Stay Free, AIDS Free, A fast-track framework for ending AIDS in children, adolescents, and young women by 2020. N.D.; <https://free.unaids.org>.
3. UNAIDS. Countdown to ZERO: global plan towards the elimination of new HIV infections among children by 2015 and keeping their mother alive. 2011.
4. World Health Organization. Global monitoring framework and strategy for the global plan towards the elimination of new HIV infections among children by 2015 and keeping their mothers alive (EMTCT). 2012.
5. Barrere B, Bartlett N, Bennett E, et al. Considerations for measuring the impact of PMTCT programmes using population-based surveys in selected high HIV prevalence countries. 2012.
6. Mary M, Penazzato M, Ciaranello A, et al. Improving estimates of children living with HIV from the Spectrum AIDS Impact Model. *AIDS (London, England)*. 2017;31(Suppl 1):S13.
7. Kalton G Sampling rare and elusive populations. New York: United Nations, Department for Economic and Social Information and Policy Analysis. 1993.
8. Wagner J, Lee S. Sampling rare populations. In: *Handbook of health survey methods*. 2015:77–106.
9. Kalton G, Anderson DW. Sampling Rare Populations. *Journal of the Royal Statistical Society: Series A (General)*. 1986;149(1):65–82.
10. Reid S Non-vertical HIV transmission to children in sub-Saharan Africa. *International journal of STD & AIDS*. 2009;20(12):820–827. [PubMed: 19948895]
11. Brown K, Williams DB, Kinchen S, et al. Status of HIV epidemic control among adolescent girls and young women aged 15–24 years—seven African Countries, 2015–2017. *Morbidity and Mortality Weekly Report*. 2018;67(1):29. [PubMed: 29329280]
12. Pufall EL, Nyamukapa C, Eaton JW, et al. HIV in children in a general population sample in East Zimbabwe: prevalence, causes and effects. *PloS one*. 2014;9(11):e113415. [PubMed: 25412251]
13. INSIDA. Inquérito Nacional de Prevalência, Riscos Comportamentais e Informação sobre o HIV e SIDA em Moçambique. 2009.
14. NASCOP. Kenya AIDS Indicator Survey 2012: Final Report. 2014.
15. Saito S, Chung H, Mahy M, et al. Pediatric HIV Treatment Gaps in 7 East and Southern African Countries: Examination of Modeled, Survey, and Routine Program Data. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2018;78:S134–S141. [PubMed: 29994836]
16. PHIA Project. PHIA Data Use Manual. 2019.
17. PHIA Project. ZAMPHIA Data Use Manual Supplement. 2019.
18. PHIA Project. ZAMPHIA Sampling and Weighting Technical Report. 2019.
19. Cochran WG. Sampling techniques. John Wiley & Sons; 2007.

20. Goga AE, Dinh T-H, Jackson DJ, et al. First population-level effectiveness evaluation of a national programme to prevent HIV transmission from mother to child, South Africa. *J Epidemiol Community Health*. 2015;69(3):240–248. [PubMed: 25371480]
21. Vrazo AC, Sullivan D, Ryan Phelps B. Eliminating Mother-to-Child Transmission of HIV by 2030: 5 Strategies to Ensure Continued Progress. *Glob Health Sci Pract*. 2018;6(2):249–256. [PubMed: 29959270]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

Number of eligible child respondents, response rates, and number of HIV positive children, by country and maternal HIV status: Population-based HIV Impact Assessment, 2015 -- 2017

	Country										Total
	Lesotho	Malawi	Namibia	Eswatini	Tanzania	Uganda	Zambia	Zimbabwe			
Total households selected	10,892	14,267	12,716	6,417	16,198	13,435	13,440	15,009		102,374	
Total child-selected households	5,443	7,132	6,358	3,206	5,340	7,926	6,733	7,511		49,649	
Total child-selected households who responded	4,445	4,580	4,059	2,599	3,336	5,809	3,561	5,149		33,538	
Total rostered children age 0-14	9,866	19,964	15,223	7,636	31,573	30,290	23,272	19,082		156,926	
HIV positive or deceased mother	2,539	2,459	1,937	2,098	2,146	2,113	2,761	2,947		19,000	
HIV unknown status mother	3,884	5,784	7,000	2,821	7,747	7,230	6,639	6,558		47,663	
HIV negative mother	3,463	11,721	6,286	2,717	21,680	20,947	13,872	9,557		90,243	
Eligible children in child-selected households	4,870	9,993	7,887	3,997	10,452	10,793	11,646	9,599		69,237	
HIV positive or deceased mother	27%	13%	13%	26%	6%	6%	12%	15%		13%	
HIV unknown status mother	38%	29%	46%	37%	25%	20%	29%	34%		31%	
HIV negative mother	35%	58%	42%	37%	69%	74%	59%	51%		57%	
Children with blood test results (response rate)	3,966 (87%)	6,166 (66%)	6,761 (91%)	3,372 (89%)	9,616 (95%)	10,345 (97%)	8,015 (79%)	7,032 (83%)		55,273 (80%)	
HIV positive or deceased mother	90%	78%	95%	92%	97%	95%	85%	88%		84%	
HIV unknown status mother	84%	47%	89%	87%	89%	96%	63%	74%		66%	
HIV negative mother	88%	72%	91%	90%	96%	98%	83%	86%		86%	
HIV prevalence among children age 0-14[†] (% , CI)	2.1 (1.5 - 2.6)	1.5 (1.1 - 1.9)	1.0 (0.8 - 1.3)	2.8 (2.2 - 3.4)	0.4 (0.2 - 0.6)	0.5 (0.3 - 0.8)	1.2 (0.9 - 1.4)	1.6 (1.2 - 2.0)		0.8 (0.7 - 0.9)	
HIV positive or deceased mother [†]	5.3 (3.8 - 6.8)	8.5 (5.9 - 11.0)	5.2 (3.6 - 6.9)	7.6 (5.7 - 9.4)	6.3 (3.7 - 8.9)	4.6 (2.8 - 6.3)	6.9 (5.2 - 8.7)	7.2 (5.5 - 9.0)		6.5 (5.6 - 7.5)	
HIV unknown status mother [†]	1.7 (0.9 - 2.5)	2.0 (0.9 - 3.2)	0.8 (0.4 - 1.2)	2.2 (1.3 - 3.1)	0.2 (0.0 - 0.3)	0.7 (0.2 - 1.2)	1.0 (0.5 - 1.6)	1.7 (0.7 - 2.8)		0.8 (0.5, 1.0)	
HIV negative mothers [†]	0.05 (0.0 - 0.1)	0.11 (0.0 - 0.2)	0.06 (0.0 - 0.1)	0.08 (0.0 - 0.2)	0.01 (0.0 - 0.03)	0.07 (0.0 - 0.2)	0.05 (0.0 - 0.1)	0.05 (0.0 - 0.1)		0.05 (0.01, 0.08)	

	Country									Total
	Lesotho	Malawi	Namibia	Eswatini	Tanzania	Uganda	Zambia	Zimbabwe		
Relative Sampling Error for overall prevalence	13%	13%	13%	11%	21%	20%	11%	12%		6.5%
Number of children tested negative for HIV	3,882	6,067	6,687	3,276	9,571	10,293	7,920	6,912		54,608
Number of HIV-positive children	84	99	74	96	45	52	95	120		665
HIV positive or deceased mother	60 (71%)	74 (75%)	49 (66%)	68 (71%)	37 (82%)	35 (67%)	74 (78%)	87 (73%)		484 (73%)
HIV unknown status mother	23 (27%)	20 (20%)	23 (31%)	27 (28%)	6 (13%)	15 (29%)	17 (18%)	31 (26%)		162 (24%)
HIV negative mother	1 (1%)	5 (5%)	2 (3%)	1 (1%)	2 (4%)	2 (4%)	4 (4%)	2 (2%)		19 (3%)

√ Weighted estimates including non-response adjustment and post-stratification, calculated with jackknife replicate weights.

Table 2:

Detailed maternal HIV status of eligible and HIV positive children, and HIV prevalence among children, by country: Population-based HIV Impact Assessment, 2015 -- 2017

	Country									Overall
	Lesotho	Malawi	Namibia	Eswatini	Tanzania	Uganda	Zambia	Zimbabwe		
Number of eligible children age 0–14 by maternal HIV status										
HIV positive mother*	942 (19%)	1,014 (10%)	740 (9%)	876 (22%)	386 (4%)	457 (4%)	1,020 (9%)	1,001 (10%)	6,436 (9%)	
HIV negative mother*	1,698 (35%)	5,841 (58%)	3,297 (42%)	1,467 (37%)	7,203 (69%)	7,985 (74%)	6,920 (59%)	4,851 (51%)	39,262 (57%)	
Unknown maternal HIV status*	2,230 (46%)	3,138 (31%)	3,850 (49%)	1,654 (41%)	2,863 (27%)	2,351 (22%)	3,706 (32%)	3,747 (39%)	23,539 (34%)	
Mother non-response/refusal [†]	312 (14%)	1,496 (48%)	627 (16%)	239 (14%)	740 (26%)	259 (11%)	1,595 (43%)	312 (14%)	6,107 (26%)	
Mother not in household [‡]	1,522 (68%)	1,377 (44%)	2,940 (76%)	1,221 (74%)	1,813 (63%)	1,838 (78%)	1,715 (46%)	1,522 (68%)	14,844 (63%)	
Mother deceased [‡]	379 (17%)	253 (8%)	250 (7%)	171 (10%)	284 (10%)	216 (9%)	384 (10%)	379 (17%)	2,407 (10%)	
Insufficient information [‡]	17 (0.8%)	12 (0.4%)	33 (0.9%)	23 (1%)	26 (0.9%)	38 (2%)	12 (0.3%)	17 (0.8%)	181 (0.8%)	
HIV positive children age 0–14 by maternal HIV status										
HIV positive mother [‡]	38 (45%)	57 (58%)	33 (45%)	52 (54%)	27 (60%)	29 (56%)	66 (70%)	52 (43%)	354 (53%)	
HIV negative mother [‡]	1 (1.2%)	5 (5.1%)	2 (2.7%)	1 (1.0%)	2 (4.4%)	2 (3.8%)	4 (4.2%)	2 (1.7%)	19 (2.9%)	
Unknown maternal HIV status [‡]	45 (54%)	37 (37%)	39 (53%)	43 (45%)	16 (36%)	21 (40%)	25 (26%)	66 (55%)	292 (49%)	
Mother non-response/refusal [‡]	1 (2%)	3 (8%)	0 (0%)	0 (0%)	2 (13%)	1 (5%)	4 (16%)	1 (2%)	12 (4%)	
Mother not in household [‡]	22 (49%)	17 (46%)	23 (59%)	27 (63%)	4 (25%)	13 (62%)	13 (52%)	30 (45%)	149 (51%)	
Mother deceased [‡]	22 (49%)	17 (46%)	16 (41%)	16 (37%)	10 (63%)	6 (29%)	8 (32%)	35 (53%)	130 (45%)	
Insufficient information [‡]	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (5%)	0 (0%)	0 (0%)	1 (0.3%)	
HIV prevalence among children age 0–14 by maternal HIV status[†] (% CI)										

	Country										Overall
	Lesotho	Malawi	Namibia	Eswatini	Tanzania	Uganda	Zambia	Zimbabwe			
HIV positive mother	4.5 (3.0 – 5.9)	8.0 (5.6 – 10.5)	5.00 (2.9 – 7.1)	7.1 (5.0 – 9.1)	8.8 (4.9 – 12.7)	5.1 (3.0 – 7.1)	7.9 (5.8 – 10.0)	6.3 (4.4 – 8.2)			7.1 (6.0 – 8.2)
HIV negative mother	0.05 (0.0 – 0.1)	0.1 (0.0 – 0.2)	0.06 (0.0 – 0.1)	0.08 (0.0 – 0.2)	0.01 (0.0 – 0.03)	0.07 (0.0 – 0.2)	0.05 (0.0 – 0.1)	0.05 (0.0 – 0.1)			0.05 (0.0 – 0.08)
Unknown maternal HIV status	2.8 (1.8 – 3.8)	3.0 (1.9 – 4.2)	1.2 (0.7 – 1.6)	3.1 (2.2 – 3.9)	0.4 (0.2 – 0.7)	1.0 (0.4 – 1.7)	1.5 (0.8 – 2.2)	2.9 (1.9 – 3.9)			1.3 (1.0 – 1.5)
Mother non-response/refusal	0.8 (0.0 – 2.4)	1.3 (0.0 – 2.9)	-	-	0.2 (0.0 – 0.6)	0.9 (0.0 – 2.9)	1.1 (0.01 – 2.2)	0.5 (0.0 – 1.6)			0.6 (0.2 – 1.1)
Mother not in household	1.8 (1.0 – 2.6)	2.3 (0.8 – 3.8)	0.9 (0.5 – 1.3)	2.4 (1.4 – 3.4)	0.1 (0.0 – 0.3)	0.7 (0.1 – 1.2)	1.0 (0.3 – 1.7)	1.8 (0.7 – 3.0)			0.8 (0.5 – 1.1)
Mother deceased	7.4 (4.0 – 10.8)	10.6 (4.8 – 16.3)	5.9 (2.7 – 9.2)	10.2 (5.3 – 15.1)	3.0 (0.9 – 5.0)	3.8 (0.4 – 7.2)	3.9 (0.8 – 7.0)	9.4 (6.1 – 12.8)			5.2 (3.8 – 6.6)
Insufficient information	-	-	-	-	-	1.1 (0.0 – 3.3)	-	-			0.7 (0.0 – 2.0)

* Percentages among all eligible children in child-selected households

† Percentages among those with unknown maternal status

‡ Percentages among all HIV positive children

§ Weighted estimates including non-response adjustment and post-stratification, calculated with jackknife replicate weights.

Table 3: Expected sample size and precision under targeted child sample designs, by country: Population-based HIV Impact Assessment, 2015 -- 2017

	Country							
	Lesotho	Malawi	Namibia	Eswatini	Tanzania	Uganda	Zambia	Zimbabwe
Overall HIV prevalence for children age 0–14 (%)	2.1%	1.5%	1.0%	2.8%	0.4%	0.5%	1.1%	1.6%
Prevalence for children with HIV positive or deceased mothers (%)	5.3%	8.5%	5.2%	7.6%	6.3%	4.6%	6.9%	7.2%
Prevalence for children with unknown HIV status mothers (%)	1.7%	2.0%	0.8%	2.1%	0.2%	0.7%	1.0%	1.7%
Prevalence for children with HIV negative mothers (%)	0.05%	0.11%	0.06%	0.08%	0.01%	0.07%	0.05%	0.05%
Proportion of all rostered children with positive or deceased mothers	26%	12%	13%	27%	7%	7%	12%	15%
Proportion of all rostered children with unknown status mothers	39%	29%	46%	37%	25%	24%	29%	34%
Proportion of all rostered children with negative mothers	35%	59%	41%	36%	69%	69%	60%	50%
Neyman allocation sampling fractions (after accounting for non-response)								
Children with positive or deceased mothers	90%	78%	95%	92%	97%	95%	85%	88%
Children with unknown status mothers	52%	39%	37%	50%	16%	38%	34%	44%
Children with negative mothers	9%	9%	10%	10%	4%	12%	7%	7%
Expected number of children tested	4,606	5,260	5,101	3,607	4,253	7,305	5,618	6,196
Scenario 1: Neyman allocation design								
Expected design effect under optimal allocation	0.69	0.64	0.63	0.71	0.35	0.54	0.50	0.61
Expected RSE	8%	9%	11%	8%	14%	12%	9%	8%
Scenario 2: Modified Neyman allocation with sample reallocated from children with negative mothers								
Expected design effect	0.60	0.57	0.52	0.62	0.25	0.58	0.40	0.51
Expected bias in estimated prevalence	-0.02%	-0.06%	-0.02%	-0.03%	-0.01%	-0.05%	-0.03%	-0.02%
Expected RRMSE	8%	8%	10%	8%	12%	12%	8%	7%
Proportional reduction in MSE, compared to PHIA subsample design with equal sample size (Scenario 3)	33%	43%	33%	27%	61%	50%	43%	46%
Scenario 3: PHIA sub-sample design pro-rated to optimal allocation sample size								
Expected RSE	12%	14%	15%	11%	31%	24%	14%	13%
Actual PHIA results using the 50% household subsample design								
Number of children tested	3,966	6,166	6,761	3,372	9,616	10,345	8,015	7,032
RSE	13%	13%	13%	11%	21%	20%	11%	12%

Sampling fractions and sample sizes for children of mothers with unknown HIV status to achieve target RSE's on pediatric prevalence based on the observed distributions of children and pediatric prevalence by mother's HIV status in phase 1 PHIA surveys

Table 4.

Target RSEs on prevalence	Required sampling fractions and corresponding sample sizes for children of unknown HIV status mothers															
	Lesotho		Malawi		Namibia		Eswatini		Tanzania		Uganda		Zambia		Zimbabwe	
%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	
0.05	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#
0.10	42%	1,618	41%	2,393	60%	4,206	29%	823	#	#	#	22%	1,430	40%	2,598	#
0.15	12%	474	12%	688	16%	1,145	9%	250	35%	2,686	35%	2,536	7%	442	12%	776
0.20	6%	238	6%	338	8%	560	4%	126	7%	570	15%	1,082	3%	222	6%	388
0.25	4%	145	4%	240	5%	337	3%	77	3%	253	9%	615	2%	135	4%	235
0.30	3%	98	2%	137	3%	227	2%	52	2%	148	6%	401	1%	92	2%	159
0.35	2%	71	2%	99	2%	163	1%	38	1%	99	4%	283	1%	66	2%	115
Number of children of deceased or HIV+ mothers to be tested		2,284	1,909	1,832	1,923	2,077	2,015	2,339	2,592							

Note

Indicates RSEs targets which cannot be achieved given the number of children in sampled households who could be tested in each PHIA country.