World Health Organization

CONCEPT NOTE

HIV DRUG RESISTANCE

SURVEILLANCE OF HIV DRUG RESISTANCE IN CHILDREN NEWLY DIAGNOSED WITH HIV BY EARLY INFANT DIAGNOSIS

DECEMBER 2017

SURVEILLANCE OF HIV DRUG RESISTANCE IN CHILDREN NEWLY DIAGNOSED WITH HIV BY EARLY INFANT DIAGNOSIS

DECEMBER 2017

World Health Organization

Surveillance of HIV drug resistance in children newly diagnosed with HIV by early infant diagnosis

ISBN 978-92-4-151254-1

Suggested citation. Surveillance of HIV drug resistance in children newly diagnosed with HIV by early infant diagnosis. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see http://www.who.int/about/licensing.

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

3TC	Lamivudine
ABC	Abacavir
AIDS	Acquired immunodeficiency syndrome
ART	Antiretroviral therapy
ARV	Antiretroviral (drugs)
ATZ/r	Atazanavir/ritonavir
CI	Confidence interval
d4T	Stavudine
DBS	Dried blood spot
DRV/r	Darunavir/ritonavir
EFV	Efavirenz
EID	Early infant diagnosis
FTC	Emtricitabine
HIV	Human immunodeficiency virus
HIVdb	Stanford HIV drug resistance algorithm
HIVDR	HIV drug resistance
IN	Integrase (region of HIV-1)
INI	Integrase inhibitor
LPV/r	Lopinavir/ritonavir
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
NVP	Nevirapine
PCR	Polymerase chain reaction
PI	Protease inhibitor
PMTCT	Prevention of mother-to-child transmission (of HIV)
PR	Protease (region of HIV-1)
RT	Reverse transcriptase (region of HIV-1)
SID	Survey identification number
TDF	Tenofovir disoproxil fumarate
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization
ZDV	Zidovudine

1. INTRODUCTION

Despite progress in scaling up interventions for the prevention of mother-to-child transmission (PMTCT) of HIV, globally an estimated 220 000 [190 000–260 000] children were newly infected with HIV in 2014. Most of these new infections occurred in sub-Saharan Africa, where more than 90% of all children infected with HIV currently live. The latest estimates indicate that, among the 2.6 million [2.4–2.8 million] children younger than 15 years living with HIV, 32% [30%–34%] were accessing treatment in 2014, up from 14% [13%–15%] in 2010.[1]

Access to antiretroviral (ARV) medicines for pregnant and breastfeeding women living with HIV has expanded. In 2013, an estimated 32% [26–36%] of mothers infected with HIV were not receiving lifelong antiretroviral therapy (ART) or prophylaxis during the breastfeeding period to reduce the risk of HIV transmission. This is a remarkable improvement from 2009, when more than 80% [79–82%] were not covered during the breastfeeding period.[2]

The continued expansion of maternal ARV drug coverage – while critical to reducing the number of new infant HIV infections – has led to increased exposure to non-nucleoside reverse transcriptase inhibitors (NNRTI). Paradoxically, this has resulted in HIV drug resistance (HIVDR) among infants and children acquiring HIV infection, despite PMTCT interventions.[3] History of NNRTI-exposure was often used as a marker to identify children who should initiate Lopinavir/ritonavir (LPV/r)-based ART. However, information about a child's ARV drug exposure is often unknown.

In 2013, the World Health Organization (WHO) recommended the use of LPV/r-based ART as the standard regimen of choice for HIV-infected children, irrespective of PMTCT exposure history.[4] Despite these guidelines, in many countries a significant proportion of children still initiate NNRTI-based regimens due to cost and feasibility issues. In these countries, an understanding of the prevalence of HIVDR among children less than 18 months of age could help accelerate the shift for this population towards LPV/r-based regimens as first-line ART.

Moreover, exposure to PMTCT ARV drugs may not be routinely recorded for children starting ART, and in many cases, previous ARV exposure is mistakenly reported as "none" or "unknown". Therefore, children who have been exposed to ARV drugs antepartum, intrapartum or postpartum may be started on an NNRTI-based regimen. Hence, it is important to evaluate the proportion of children whose PMTCT history is reported as "none" or "unknown" having mutations associated with resistance, which may affect their treatment outcome. This is particularly relevant in countries considering the introduction of protease inhibitor (PI)-sparing strategies once viral load suppression is sustained, and in countries where NNRTI-based regimens are used as second-line.

1. UNAIDS. Fact sheet 2015. Available at: http://www.unaids.org/en/resources/campaigns/HowAIDSchangedeverything/factsheet

2. UNAIDS. The GAP. UNAIDS. Geneva, Switzerland. 2014. Available at: http://www.unaids.org/sites/default/files/en/media/unaids/contentassets/documents/unaidspublication/2014/UNAIDS_Gap_report_en.pdf

3. Kuhn L, Hunt G, Technau K-G, et al. Drug resistance among newly diagnosed HIV-infected children in the era of more efficacious antiretroviral prophylaxis. AIDS. 2014;28(11):1673–1678.

4. WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection. WHO. Geneva, Switzerland. 2013. Available at: http://www.who.int/hiv/pub/guidelines/arv2013/en/

2. SURVEY PURPOSE

This concept note describes the methods used to assess the prevalence of any HIVDR and HIVDR by PMTCT exposure among children less than 18 months of age using remnant dried blood spot (DBS) specimens from early infant diagnosis (EID) over a 12-month period. Data on HIVDR and the prevalence of PMTCT exposure in this target population can provide critical information to support optimal choice of first- and second-line ART regimens. Survey limitations and potential biases are discussed in Box 1.

Box 1: Survey limitations and potential biases

The methodology presented in this concept note relies on remnant DBS specimens used for EID to assess the prevalence of HIVDR amongst treatment-naïve children less than 18 months. While this approach has important operational and practical advantages, its results must be interpreted in light of its limitations.

In particular, survey results may not necessarily represent all children less than 18 months of age infected or diagnosed with HIV-1 in the country. If EID coverage is low, many children may go undiagnosed, and survey results may not necessarily reflect the total population of children less than 18 months newly infected with HIV.

The type of sites contributing specimens to the survey may also introduce potential biases in survey results. If PMTCT sites contribute most diagnostic specimens, children with "no" or "unknown" ARV exposure may not be well represented in the survey. Alternatively, if non-PMTCT sites contribute most diagnostic specimens, children with recorded ARV exposure may not be well represented.

3. SURVEY OUTCOMES

The survey has nine main outcomes. The first six outcomes (1a, 1b, 2a, 2b, 3a and 3b) provide measures of HIVDR, while the remaining three (4a, 4b and 4c) describe the prevalence of "yes", "no" and "unknown" PMTCT exposure in the target population. These outcomes are summarized in Table 3.1.

Outcomes 1a and 1b summarize the prevalence of HIVDR regardless of PMTCT exposure. Outcomes 2a, 2b, 3a and 3b summarize the prevalence of HIVDR stratified by PMTCT exposure.

1a. Prevalence of any HIVDR[1,2] among all treatment-naïve children less than 18 months of age newly diagnosed with HIV, regardless of PMTCT exposure.

1b. Prevalence of HIVDR to NNRTI (NVP or EFV) among all treatment-naïve children less than 18 months of age newly diagnosed with HIV, regardless of PMTCT exposure.

2a. Prevalence of any HIVDR among treatment-naïve children less than 18 months of age newly diagnosed with HIV with known PMTCT exposure (maternal only, neonatal only or both).

2b. Prevalence of HIVDR to NNRTI (NVP or EFV) among treatment-naïve children less than 18 months of age newly diagnosed with HIV with known PMTCT exposure (maternal only, neonatal only or both).

3a. Prevalence of any HIVDR among treatment-naïve children less than 18 months of age newly diagnosed with HIV with no or unknown PMTCT exposure.[3]

3b. Prevalence of HIVDR to NNRTI (NVP or EFV) among treatment-naïve children less than 18 months of age newly diagnosed with HIV with no or unknown PMTCT exposure.[3]

These outcomes must be calculated taking into account *observed population sizes*[4] and the number of *case specimens*[5]

1. Any HIVDR is defined with respect to one or more of the following drugs or drug classes: NVP, EFV, any N(t)RTI, DRV/r, LPV/r or ATV/r. Sequences classified as low-level, intermediate or high-level resistance according to the Stanford HIVdb are aggregated as "HIVDR".

2. In countries opting to genotype the IN region of HIV-1, detected INI resistance is excluded when estimating the prevalence of any HIVDR.

3. In country-level analyses HIVDR outcomes in the "unknown" and "no exposure" category are aggregated due to anticipated small sample sizes; however, global aggregate analyses may have sufficient power to assess differences amongst those with no versus unknown exposure.

4. The *observed population size* is the number of stored specimens available that were collected and stored during the time period of interest.

5. In this concept note, *case specimen* is defined as a remnant EID specimen obtained from a child meeting the inclusion/exclusion criteria described in Section 4.3.

with sequences genotyped at each laboratory. The analysis will account for these elements through adjustments of the survey weights (an example of a data analysis plan is provided in the Annex, Section 7.1, and additional technical background is available in the Statistical Appendix, Section 8).

Outcomes 4a, 4b and 4c summarize the prevalence of PMTCT exposure among treatment-naive children less than 18 months of age.

4a. Proportion of treatment-naïve children less than 18 months of age newly diagnosed with HIV with PMTCT exposure (maternal only, neonatal only or both).

4b. Proportion of treatment-naïve children less than 18 months of age newly diagnosed with HIV without PMTCT exposure.

4c. Proportion of treatment-naïve children less than 18 months of age newly diagnosed with HIV with unknown PMTCT exposure.

Outcomes 4a, 4b and 4c must be calculated taking into account *observed population sizes* and observed number of case specimens with available PMTCT exposure information by laboratory. The analysis will account for these elements through adjustments of the survey weights.

Table 3.1: Summary of survey outcomes

	Outcome	Target population
HIVDR	Outcome 1a. Prevalence of any HIVDR	Treatment-naive children < 18 months of age, *regardless* of PMTCT exposure
	Outcome 1b. Prevalence of HIVDR to NNRTI	
	Outcome 2a. Prevalence of any HIVDR	Treatment-naïve children < 18 months of age with *known* PMTCT exposure
	Outcome 2b. Prevalence of HIVDR to NNRTI	
	Outcome 3a. Prevalence of any HIVDR	Treatment-naïve children < 18 months of age with *no or unknown* PMTCT exposure
	Outcome 3b. Prevalence of HIVDR to NNRTI	
Prevalence of PMTCT exposure	Outcome 4a. Proportion of treatment-naive children < 18 months of age newly diagnosed with HIV *with* PMTCT exposure	
	Outcome 4b. Proportion of treatment-naive children < 18 months of age newly diagnosed with HIV *without* PMTCT exposure	
	Outcome 4c. Proportion of treatment-naive children < 18 months of age newly diagnosed with HIV *with unknown* PMTCT exposure	

4. OVERVIEW OF METHODS

4.1 General approach

The methodology is a retrospective survey of stored remnant DBS collected for paediatric polymerase chain reaction (PCR) diagnosis of HIV during a recent 12-month period. HIV molecular diagnostic methods involving DBS are being scaled up in an increasing number of low- and middle-income countries. In most settings, three to five DBS are collected from each child for PCR testing and other purposes, and remnant specimens are likely to be available for HIVDR surveillance purposes.

All laboratories where EID is performed in the country will contribute case specimens to the survey, with the number of case specimens contributed per laboratory depending on their respective size (see Section 4.5 for additional details). The reverse transcriptase (RT), protease (PR) and integrase (IN) regions of the HIV-1 *pol* gene will be sequenced using standard sequencing methods.[1] Amplification and sequencing should be performed at laboratories designated by WHO for genotyping

1. INI are infrequently used in resource-limited settings. Genotyping of the IN region should only be considered if INI containing regimens are used as part of national treatment guidelines. At present, it is anticipated that most countries will not opt to genotype this region.

using DBS as a specimen type (see Section 4.6.2). Demographic information and clinical data, including PMTCT regimen exposure, will be abstracted from laboratory requisition forms, with no participant-level identifying information recorded.

The analysis consists of calculating the point prevalence estimates and 95% confidence intervals (CIs) for the outcomes described in Section 3. HIVDR will be determined using the Stanford HIVDR algorithm (HIVdb).[1] For additional analyses when sample sizes within a country are insufficient, survey results can be aggregated in multi-country analyses to achieve greater precision.

4.2 EID laboratory inclusion

All EID laboratories in the country should contribute case specimens to the survey, but only a subset of available case specimens will be selected from each laboratory (discussed in Section 4.5).

Some countries may have EID laboratories that are difficult to access for a variety of reasons, such as logistical complications, political constraints or geographical remoteness. Although not advisable, countries may consider excluding some of these laboratories from the survey. In general, if less than 10% of eligible EID specimens are stored at these laboratories, countries may choose to exclude them. The 10% threshold seeks to limit the potential bias that such exclusion may introduce in the final results. In this case, exclusion of a laboratory should be done a priori (and not after the lab has been included in the sample size calculations). A list of all excluded laboratories and reasons for their exclusion should be recorded in any resulting technical report. On the other hand, if more than 10% of eligible EID specimens are stored at these laboratories, it is not advisable to exclude them from the survey. In general, if the excluded case specimens from these laboratories have a different prevalence of HIVDR than the observed case specimens from the included laboratories, the national prevalence estimate will be biased.

4.3 Specimen eligibility criteria

Inclusion criteria

A case specimen will be considered eligible under the following conditions:

1. Child is less than 18 months of age;
2. DBS tested HIV-positive by PCR.[2]

Exclusion criteria

1. Child is 18 months of age or older;
2. Child is receiving three or more ARV drugs for the purpose of HIV treatment (rather than prophylaxis to prevent HIV infection) at time of blood draw.

4.4 Defining the survey sample size

Assumptions

A number of key model assumptions affect the required survey sample size for Outcomes 1a/b, 2a/b and 3a/b. These assumptions are summarized in Table 4.1.

Table 4.1: Key model assumptions to calculate the sample size

Assumptions	Proposed values
Expected prevalence of drug resistance	50%
Expected genotyping failure rate	20%
Expected proportion of EID case specimens with known PMTCT exposure	80%
Desired CI half-width for Outcomes 1a/b	±5%
Maximum desired CI half-width for Outcomes 2a/b and 3a/b	±11%

Recent evidence has uncovered very high levels of HIVDR among children less than 18 months of age newly diagnosed with HIV. For example, in 2013 in Zimbabwe, 62.5% of case specimens sampled had NNRTI-related resistance.[3] Therefore, the expected prevalence of drug resistance is conservatively assumed to be 50%. This assumption will yield the largest sample size and hence the most precise CIs. Given the use of routinely collected DBS specimens that may have been suboptimally handled and/or stored, the expected genotyping failure rate is assumed to be 20%.

1. Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin Infect Dis*. 2006;42(11):1608–18. Epub 2006 Apr 28.

2. If more than one specimen is available for the same child, the specimen that tested positive first by PCR should be used.

3. Chakanyuka-Musanhu CC, Penazzato M, Apollo T et al. WHO HIV drug resistance surveillance in children less than 18 months newly diagnosed with HIV in Zimbabwe. 2013. Poster number TUPE278, presented at the International AIDS Conference, 2013. Available at: http://www.who.int/hiv/pub/posters/iaspost_n15_dr/en/

The survey is designed to have the greatest precision for Outcomes 1a/b (±5%) and to place an upper limit on the CI half-width of Outcomes 2a/b and 3a/b, the magnitude of which will depend on what proportion of the population has known PMTCT exposure. The expected proportion of EID case specimens with PMTCT exposure is assumed to be 80%.[1] As a result, it is assumed that the proportion of EID case specimens with no or unknown PMTCT exposure is 20%. A maximum CI half-width of ±11% is suggested as an appropriate compromise between feasibility and precision; this limit is reached when the known versus no or unknown PMTCT exposure breakdown is 80%/20%, although the CI is expected to be narrower (more precise) when the breakdown is closer to 50%/50%.

Under the assumptions outlined in Table 4.1, the recommended sample size for a country is 500 in total (more details are available in Section 8). If this sample size exceeds the number of eligible case specimens in the country, the country should perform a census of all available case specimens.

In practice, it is expected that there will be most variability around Outcomes 3a/b because there may be fewer case specimens with no or unknown PMTCT exposure. Thus, superior precision is expected for Outcomes 2a/b (approximately ±5–6%) because more case specimens will have known PMTCT exposure, resulting in a larger sample size. The predicted CI half-widths for a sample size of 500, given different population breakdowns of PMTCT exposure, are outlined in Table 4.2. The CIs for Outcomes 2a/b and 3a/b only approach ±11% when the breakdown of known versus no or unknown PMTCT exposure is far from 50%/50%.

Table 4.2: Predicted CI half-widths for Outcomes 1a/b, 2a/b and 3a/b assuming different proportions with known versus no or unknown PMTCT exposure. Assumes 500 samples collected with 20% laboratory failure and 50% prevalence of HIVDR

Proportion with known exposure	Predicted CI (Outcomes 2a/b)	Proportion with no or unknown exposure	Predicted CI (Outcomes 3a/b)	Predicted overall CI (Outcomes 1a/b)
80%	±5.5%	20%	±11.0%	±4.9%
70%	±5.9%	30%	±8.9%	±4.9%
60%	±6.3%	40%	±7.7%	±4.9%
50%	±6.9%	50%	±6.9%	±4.9%
40%	±7.7%	60%	±6.3%	±4.9%
30%	±8.9%	70%	±5.9%	±4.9%
20%	±11.0%	80%	±5.5%	±4.9%

4.5 How to allocate the sample size to each EID laboratory

All available case specimens from the selected 12-month period that meet the eligibility criteria should be available for sampling in the survey. If this is not the case, survey results may be biased.

If there is only one EID laboratory in the country, all case specimens will be sampled from that laboratory. If there is more than one EID laboratory in the country, the overall sample size will be distributed across these laboratories in a manner proportional to their size. To determine the appropriate distribution of sample sizes, the country must first list all laboratories to be included in the survey. It must then determine the number of eligible case specimens from each laboratory. The number of eligible case specimens is a count of remnant EID case specimens from each laboratory during the target 12-month survey period. If 50% of EID case specimens are at a particular laboratory, then 50% of the survey sample size should be collected at this laboratory. It may be necessary to round the sample size to the nearest whole number. See Box 2 for an example.

Once the appropriate sample size per laboratory has been determined, the laboratory must randomly sample case specimens for inclusion in the survey. It is recommended that case specimens are selected using systematic sampling. A set of specimens is identified at each laboratory, based on the date of specimen collection being within the target 12-month survey period. The size of this set of specimens is counted and stored, as it will be necessary to conduct the analysis. If there are 1000 specimens in the set, and the laboratory must sample 100 specimens, systematic sampling dictates that the laboratory must sample every 1000/100 = 10 specimens (if this number includes a decimal, it can be rounded down to the nearest whole number). The laboratory selects a specimen from the first 10 specimens in the set as a random starting point (e.g. the fifth specimen), then samples every 10th specimen (e.g. the 15th specimen, the 25th specimen, etc.) until the sample size is achieved. No case specimen should be sampled more than once.

1. Penazzato M. WHO HIV drug resistance surveillance in children less than 18 months old newly diagnosed with HIV: results from Swaziland and Zimbabwe. 5th International Workshop on HIV Pediatrics, Kuala Lumpur, Malaysia: 28–29 June 2013 (Abstract O_2012AB).

Box 2: Allocating the sample size proportionally to the size of the laboratories

Suppose there are two laboratories in Country Y. Laboratory A has remnant case specimens from 3000 newly infected children in the target year, and Laboratory B has remnant case specimens from 2000 newly infected children in the target year, making a total of 5000. Laboratory A will be allocated 3000/5000 = 60% of the sample size, and Laboratory B will be allocated 2000/5000 = 40% of the sample size. As the overall sample size is 500, Laboratory A will genotype 300 case specimens (60% of 500), and Laboratory B will genotype 200 case specimens (40% of 500).

4.6 Laboratory methods

4.6.1 Specimen collection, handling, processing and tracking

At least one viable remnant DBS not required for clinical testing or quality assurance must be available (between two and four DBS are optimal.) If DBS for PCR is collected from a child at different time points, these should be clearly labelled with unique identifiers so that the child is not counted more than once. If more than one DBS is available for the same child, the first one should be used.

The survey will use remnant DBS available after all diagnostic, clinical and quality assurance tests have been performed. When collecting DBS for HIV diagnosis by PCR, national or site-based guidance should be followed. To be suitable for genotyping, DBS should be handled, transported and stored according to the WHO recommendations for HIVDR testing on DBS.[1] For more detailed information on processing, handling, storage, transport within the country, and shipment outside the country, refer to the *WHO manual for HIVDR testing using DBS specimens* (http://www.who.int/hiv/topics/drugresistance/dbs_protocol.pdf).

4.6.2 HIVDR genotyping and quality assurance of sequences

Specimens collected for HIVDR surveillance should be tested in WHO-designated HIVDR genotyping laboratories. As members of the WHO HIVResNet Laboratory Network, these laboratories undergo a rigorous inspection process and participate in annual proficiency panel testing. Designated laboratories perform extensive quality assurance of sequences and should follow the WHO *Laboratory standard operating procedures for post-testing quality assurance of HIVDR genotyping*. Use of WHO-designated laboratories promotes quality-assured results for the purpose of public health surveillance. If a country does not have a WHO-designated laboratory for HIVDR testing, it is encouraged to send specimens to a WHO-designated regional or specialized laboratory. A list of WHO-designated laboratories may be found on the WHO HIVDR webpage.[2]

5. IMPLEMENTATION CONSIDERATIONS

5.1 Convention for assigning survey identification numbers

Once case specimens have been selected for genotyping, they must be assigned a survey identification number (SID), or unique survey ID. A logbook should be kept matching the SID to the case specimen's original identification number. The SID will be used to identify the patient, as well as the sequence generated by the genotyping assay. It is composed of the following five elements, delimited by a dash character ("-"):

1. Country abbreviation: the standard three-letter abbreviation, as defined by the International Organization for Standardization;[3]

1. http://www.who.int/hiv/topics/drugresistance/dbs_protocol.pdf
2. http://www.who.int/hiv/topics/drugresistance/AccreditedLabsFeb2014.pdf
3. http://www.worldatlas.com/aatlas/ctycodes.htm

2. Survey type: INF;
3. Year when the first specimen was collected;
4. Laboratory abbreviation: a three-letter abbreviation for the laboratory performing the EID testing, unique within the country (by default, the first three letters of the laboratory name, unless this is not unique);
5. Four-digit unique patient number: a consecutive unique patient number assigned to a case specimen sampled at a laboratory for this survey.

For example, if the "University HIV Laboratory" participated in a national survey of HIVDR among infants less than 18 months in South Africa in 2016, the first case specimen's SID would be: ZAF-INF-2016-UNI-0001.

5.2 Data abstraction

Information should be abstracted from the laboratory forms accompanying the DBS into survey data abstraction forms. Patient data should be abstracted from laboratory requisition forms that accompany DBS specimens, which may be supplemented by records or registers maintained at sites where blood is drawn. The same unique SID should be assigned to the data abstraction form and the DBS specimen.

5.3 List of variables to be collected

5.3.1 Required patient-level variables

1. EID laboratory ID;
2. Survey ID (as specified in Section 5.1);
3. Date of birth; if not available, age of child in months on date of DBS collection;
4. Date of DBS collection;
5. Mother exposed to ARV drug(s) during pregnancy and/or breastfeeding? (This includes exposure for the purpose of treatment for her health or for PMTCT): (yes/no/unknown);
6. Child received postnatal ARV drug prophylaxis? (yes/no/unknown);
7. RT region of the HIV-1 pol gene successfully sequenced? (yes/no);[1]
8. PR region of the HIV-1 pol gene successfully sequenced? (yes/no);[1]
9. IN region of the HIV-1 pol gene successfully sequenced? (yes/no/not applicable);[2]
10. Date of genotyping (DD/MM/YY);
11. Drug resistance for the following drugs: NVP, EFV, ETR, RPV, d4T, ZDV, TDF, ABC, 3TC/FTC, DRV/r, LPV/r, ATZ/r. [3]

5.3.2 Required laboratory-level variables

1. Name of laboratory;
2. Number of EID case specimens available at the laboratory (size of eligible population during a defined 12-month period);
3. Number of EID case specimens sampled at the laboratory;
4. Number of EID case specimens successfully genotyped;
5. Screening start date (date the first DBS specimen was sampled from the first EID laboratory);
6. Screening complete date (date the last DBS specimen was sampled from the last EID laboratory).

---

1. A specimen is considered to be successfully sequenced only after it passes the appropriate quality assurance recommended by WHO.

2. INI are infrequently used in resource-limited settings. Genotyping of the IN region should only be considered if INI containing regimens are used as part of national treatment guidelines. At present, it is anticipated that most countries will not opt to genotype this region. If IN region is sequenced, resistance to INI is reported.

3. For this survey, the Stanford HIVdb is used to classify HIVDR. This algorithm classifies HIVDR into five categories: susceptible, potential low-level, low-level, intermediate or high-level drug resistance. Sequences classified as susceptible and potential low-level resistance are considered to have no HIVDR.

6. DATA ANALYSIS

6.1 Definition of HIVDR

For this survey, the Stanford HIVdb is used to classify HIVDR. This algorithm classifies HIVDR in five levels: susceptible, potential low-level, low-level, intermediate, or high-level drug resistance.

Outcomes 1a, 2a and 3a measure the prevalence of any HIVDR, defined as low-level, intermediate or high-level resistance (according to the Stanford HIVdb) to one or more of the following drugs or drug classes: NVP, EFV, any N(t)RTI, DRV/r, LPV/r or ATV/r. Sequences classified as susceptible and potential low-level resistance are considered to have no HIVDR.

Outcomes 1b, 2b and 3b measure the prevalence of NNRTI drug resistance, defined as low-level, intermediate or high-level resistance (according to the Stanford HIVdb) to NVP, EFV or both. Sequences classified as susceptible and potential low-level resistance are considered to have no HIVDR.

6.2 Data analysis plan

Once data have been abstracted and genotyping is complete, point prevalence estimates will be calculated for: (1) any HIVDR; (2) HIVDR in children with known PMTCT exposure; (3) HIVDR in children with no or unknown PMTCT exposure; and (4) the prevalence of each category of PMTCT exposure (yes/no/unknown), along with their respective CIs. Classification of PMTCT exposure based on maternal and neonatal exposure is described in Table 6.1. Data are weighted taking into account laboratory sizes and observed genotyping failure rates. Guidance on data analysis is provided in the Annex, Section 7. 1. Additional technical background can be found in the Statistical Appendix, Section 8.

Table 6.1: Classification of PMTCT exposure based on maternal and neonatal exposure

		Neonatal exposure		
		Yes	No	Unknown
Maternal exposure	Yes	Yes	Yes	Yes
	No	Yes	No	Unknown
	Unknown	Yes	Unknown	Unknown

The survey is not powered to generate precise estimates of resistance among children less than 18 months with different types of PMTCT exposure (maternal only, neonatal only, both, unknown). However, data from children less than 18 months with PMTCT exposure can be aggregated across surveys to obtain regional and global estimates with an acceptable CI.

To facilitate countries' reporting of survey data to WHO, a standardized Excel-based reporting template will be available for download on the WHO HIVDR website at http://www.who.int/hiv/topics/drugresistance/en/index.html

7. ANNEX

7.1 Data analysis plan

It is recommended that data entry be conducted in Excel, and data analysis in Stata. Instructions for data analysis in Stata are provided in this section. Alternative statistical packages can be used to perform data analysis, provided they properly adjust for survey weights and stratification.

An example of a survey is provided in Box 2 (Section 4.5). In this example, there are two laboratories (Laboratory A and Laboratory B) that process EID case specimens in the country. The overall sample size is 500. From Laboratory A, out of 3000 eligible case specimens, 300 are selected for genotyping using systematic sampling (every 10th specimen selected). From Laboratory B, out of 2000 eligible case specimens, 200 are selected for genotyping using systematic sampling (every 10th specimen selected).

Step I: Create a table summarizing the necessary information for each case specimen sampled

In Excel, create a spreadsheet summarizing the necessary information for each case specimen selected (see Table 7.1 for an example).

1. List the unique case specimen ID (SID) in a column labelled `ID`;
2. List the three-character laboratory ID in a column labelled `LAB_ID`;[1]
3. List a number indicating the case specimen's PMTCT exposure (0 if no known exposure; 1 if known exposure; 2 if unknown exposure) in a column labelled `PMTCT_EXPOSURE_RF`. The strategy for classifying PMTCT exposure based on maternal and neonatal exposure is described in Table 6.1;
4. List a binary variable indicating whether a case specimen was successfully genotyped in a column labelled `GENOTYPED_BN` (1 if successfully genotyped and results available; 0 otherwise);
5. List a binary variable indicating whether the case specimen had any detected HIVDR in a column labelled `ANY_HIVDR_BN` (1 if HIVDR[2]; 0 if no HIVDR; blank if case specimen not successfully genotyped or results unavailable);
6. List a binary variable indicating whether the case specimen had any detected NNRTI HIVDR in a column labelled `NNRTI_HIVDR_BN` (1 if NNRTI HIVDR; 0 if no NNRTI HIVDR; blank if case specimen not successfully genotyped or results unavailable);
7. Save data in a spreadsheet, such as `INF_DATA.xlsx`.

Table 7.1: Example of case specimen data

ID	LAB_ID	PMTCT_ EXPOSURE_RF	GENOTYPED_BN	ANY_HIVDR_BN	NNRTI_HIVDR_ BN
XYZ-INF-2016-AAA-0001	AAA	1	1	1	1
XYZ-INF-2016-AAA-0002	AAA	0	1	1	0
….					
XYZ-INF-2016-BBB-0001	BBB	1	0		
XYZ-INF-2016-BBB-0002	BBB	2	1	0	0

1. It is recommended that the laboratory ID correspond exactly to the three-letter site code described in Section 5.1.
2. Refer to Section 6.1.

In the above sample table:

1. Case specimen XYZ-INF-2016-AAA-0001 (from Laboratory A) had known PMTCT exposure and evidence of NNRTI HIVDR;
2. Case specimen XYZ-INF-2016-AAA-0002 (from Laboratory A) had no known PMTCT exposure and evidence of HIVDR but no NNRTI HIVDR;
3. Case specimen XYZ-INF-2016-BBB-0001 (from Laboratory B) had known PMTCT exposure but the specimen was not successfully genotyped;
4. Case specimen XYZ-INF-2016-BBB-0002 (from Laboratory B) had unknown PMTCT exposure and no evidence of HIVDR.

Step II: Import data into Stata

1. Import patient data using the import data option (*File/Import/Excel Spreadsheet*).[1] Use the Browse button to identify the spreadsheet. Select the option to Import the first row as variable names. Change the variable case to upper to preserve variable names;
2. Save patient data as a `.dta` file using the save option (*File/Save*). In this example, we save the data as `INF_DATA.dta`. Press Yes to overwrite data currently in memory.

Step III: Create survey weights and other necessary variables

1. Count number of case specimens sampled by lab. These are case specimens that were selected by systematic sampling and for which demographic data were abstracted. In the command line, type:

```
bysort LAB_ID : egen N_SPECIMENS_SAMPLED = _N
```

2. Count number of specimens genotyped by lab. In the command line, type:

```
bysort LAB_ID : egen N_SPECIMENS_GENOTYPED = total(GENOTYPED_BN == 1)
```

3. Input size of each lab into Stata. We use the hypothetical data from our example. In the command line, type:

```
generate LAB_SIZE = .

replace LAB_SIZE = 3000 if LAB_ID == "AAA"

replace LAB_SIZE = 2000 if LAB_ID == "BBB"
```

4. Create survey weight for Outcomes 1, 2 and 3. In the command line, type:

```
generate OUTCOME123_WT = (LAB_SIZE/N_SPECIMENS_GENOTYPED)
```

5. Create survey weight for Outcome 4. In the command line, type:

```
generate OUTCOME4_WT = (LAB_SIZE/N_SPECIMENS_SAMPLED)
```

6. Create a variable to be used for reporting results for global aggregation. In the command line, type:

```
generate POP_SIZE = 1
```

1. For this section, formatting indicates the following: Stata/Dropdown Menu/Directions, `variable_name` or `type into command line`, and Stata option.

Step IV: Declare survey design and analyse data

1. Declare survey design for Outcomes 1, 2 and 3 (*Statistics/Survey data analysis/Setup & utilities/Declare survey design for dataset*):
   a. In the *Main tab*, change *Number of stages* to 1;
   b. Select `ID` as the Stage 1: *Sampling units;*
   c. Select `LAB_ID` as the Stage 1: *Strata*;
   d. In the *Weights* tab, select `OUTCOME123_WT` as the *Sampling weight variable;*
   e. In the *SE* tab, select *Center at the grand mean* for Strata with a single sampling unit. *Press OK.*

2. Analyse Outcomes 1a/b (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*):
   a. In the *Model* tab, select or type `ANY_HIVDR_BN NNRTI_HIVDR_BN` as the *Variable(s)*. *Press OK;*
   b. The total number of genotyped specimens is labelled as the *Number of obs*. Here, it is 431 observations;
   c. The point estimate, standard error and 95% CI for the prevalence of any HIVDR and NNRTI HIVDR among all treatment-naïve children less than 18 months of age newly diagnosed with HIV, regardless of PMTCT exposure, are located in the rows labelled 1. Here, the prevalence of any HIVDR is 56.8% (95% CI: 52.1, 61.5%), and the prevalence of NNRTI HIVDR is 46.6% (95% CI: 41.9, 51.4%).

```
. svy linearized : proportion ANY_HIVDR_BN NNRTI_HIVDR_BN

Survey: Proportion estimation

Number of strata =       2              Number of obs    =      431
Number of PSUs   =     431              Population size  =     5000
                                        Design df        =      429

--------------------------------------------------------------------
                |             Linearized
                | Proportion   Std. Err.     [95% Conf. Interval]
----------------+---------------------------------------------------
ANY_HIVDR_BN    |
             0  |   .4315231   .0239067       .3852821    .4789894
             1  |   .5684769   .0239067       .5210106    .6147179
----------------+---------------------------------------------------
NNRTI_HIVDR_BN  |
             0  |    .533647   .0240854        .486146    .5805456
             1  |    .466353   .0240854       .4194544     .513854
--------------------------------------------------------------------
```

3. Analyse Outcomes 2a/b (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*):

   a. In the *Model* tab, select or type `ANY_HIVDR_BN NNRTI_HIVDR_BN` as the *Variable(s)*;

   b. In the if/in/over tab, type `PMTCT_EXPOSURE_RF==1` (selects only case specimens with known PMTCT exposure) into the If: (*expression*) box. Press OK;

   c. The number of observations used to calculate this outcome is labelled *Subpop. no. obs*. Here it is 274 observations. Disregard the value labelled *Number of obs*.;

   d. The point estimate, standard error and 95% CI for the prevalence of any HIVDR and NNRTI HIVDR among treatment-naïve children less than 18 months of age newly diagnosed with HIV with known PMTCT exposure are located in the rows labelled 1. Here, the prevalence of any HIVDR is 64.2% (95% CI: 58.4, 69.7%), and the prevalence of NNRTI HIVDR is 51.1% (95% CI: 45.2, 57.0%).

```
. svy linearized, subpop(if PMTCT_EXPOSURE_RF == 1) : proportion ANY_HIVDR_BN
NNRTI_HIVDR_BN

Survey: Proportion estimation

Number of strata =       2            Number of obs    =     454
Number of PSUs   =     454            Population size  = 5266.77
                                      Subpop. no. obs  =     274
                                      Subpop. size     = 3178.92
                                      Design df        =     452

--------------------------------------------------------------
                |             Linearized
                | Proportion   Std. Err.     [95% Conf. Interval]
----------------+---------------------------------------------
ANY_HIVDR_BN    |
              0 |    .357747   .0289902       .3029865    .4164907
              1 |    .642253   .0289902       .5835093    .6970135
----------------+---------------------------------------------
NNRTI_HIVDR_BN  |
              0 |   .4891746   .0301966       .4302595    .5483919
              1 |   .5108254   .0301966       .4516081    .5697405
--------------------------------------------------------------
```

4. Analyse Outcomes 3a/b (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*):

   a. In the *Model* tab, select or type `ANY_HIVDR_BN NNRTI_HIVDR_BN` as the *Variable(s);*

   b. In the *if/in/over* tab, type `PMTCT_EXPOSURE_RF != 1` (selects only case specimens with no known or unknown PMTCT exposure) into the *If: (expression)* box. *Press OK*;

c. The number of observations used to calculate this outcome is labelled *Subpop. no. obs.* Here it is 157 observations. Disregard the value labelled *Number of obs*.;

d. The point estimate, standard error and 95% CI for the prevalence of any HIVDR and NNRTI HIVDR among treatment-naïve children less than 18 months of age newly diagnosed with HIV with no or unknown PMTCT exposure are located in the rows labelled *1*. Here, the prevalence of any HIVDR is 44.0% (95% CI: 36.4, 51.8%), and the prevalence of NNRTI HIVDR is 38.9% (95% CI: 31.6, 46.7%).

```
. svy linearized, subpop(if PMTCT_EXPOSURE_RF != 1 ) : proportion ANY_HIVDR_
BN NNRTI_HIVDR_BN


Survey: Proportion estimation

Number of strata =       2          Number of obs    =     477
Number of PSUs   =     477          Population size  = 5533.74
                                    Subpop. no. obs  =     157
                                    Subpop. size     = 1821.08
                                    Design df        =     475

--------------------------------------------------------------
             |             Linearized
             | Proportion   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
ANY_HIVDR_BN |
           0 |   .5603086   .0395656       .4817188    .6359865
           1 |   .4396914   .0395656       .3640135    .5182812
-------------+------------------------------------------------
NNRTI_HIVDR_BN |
           0 |   .6112793    .038873       .5327612    .6844185
           1 |   .3887207    .038873       .3155815    .4672388
--------------------------------------------------------------
```

5. Declare survey design for Outcome 4 (*Statistics/Survey data analysis/Setup & utilities/Declare survey design for dataset*):

a. In the *Main and SE* tabs, select the same options described in Step 1;

b. In the *Weights* tab, select `OUTCOME4_WT` as the *Sampling weight variable*. *Press OK*.

6. Analyse Outcomes 4a/b/c (*Statistics/Survey data analysis/Tables/One-way tables*):

a. In the *Model* tab, select `PMTCT_EXPOSURE_RF` as the *Categorical variable*;

b. In the *Table items* table, check the boxes for *Standard errors* and *Confidence intervals*. *Press OK*;

c. The number of observations used to calculate these outcomes is labelled *Number of obs*. Here it is 500 observations;

d. The point estimate, standard error and 95% CI for Outcome 4a (the proportion of treatment-naïve children less than 18 months of age newly diagnosed with HIV with known PMTCT exposure) are located in the row labelled *1*. Here, the estimate is 64.0% (95% CI: 59.7, 68.1%);

e. The point estimate, standard error and 95% CI for Outcome 4b (the proportion of treatment-naïve children less than 18 months of age newly diagnosed with HIV with no known PMTCT exposure) are located in the row labelled *0*. Here, the estimate is 20.8% (95% CI: 17.5, 24.6%);

f. The point estimate, standard error and 95% CI for Outcome 4c (the proportion of treatment-naïve children less than 18 months of age newly diagnosed with HIV with unknown PMTCT exposure) are located in the row labelled *2*. Here, the estimate is 15.2% (95% CI: 12.3, 18.6%).

```
. svy linearized : tabulate PMTCT_EXPOSURE_RF, se ci

Number of strata   =  2              Number of obs      =    500
Number of PSUs     =  500            Population size    =   5000
                                     Design df          =    498

-------------------------------------------------------------
PMTCT_PRI |
OR_EXPOSU |
RE_RF     | proportions           se           lb           ub
----------+--------------------------------------------------
        0 |        .208        .0182        .1745         .246
        1 |         .64        .0215        .5968        .6811
        2 |        .152        .0161         .123        .1864
          |
    Total |           1
-------------------------------------------------------------

  Key:  proportions  =  cell proportions
        se           =  linearized standard errors of cell proportions
        lb           =  lower 95% confidence bounds for cell proportions
        ub           =  upper 95% confidence bounds for cell proportions
```

7. Analyse data for aggregate reporting to WHO, as described in Section 6.2. Example shown is for Outcomes 2a/b (the prevalence of any HIVDR and NNRTI HIVDR among treatment-naïve children less than 18 months of age newly diagnosed with HIV with known PMTCT exposure):

a. Repeat Step 1 to declare the survey design used for Outcomes 1, 2 and 3;

b. Repeat Step 3 to analyse Outcomes 2a/b;

c. From the output for Outcomes 2a/b, store the following data elements:

i. The number of observations used to construct the estimate: for Outcomes 2 and 3, it is labelled *Subpop. no. obs*; for Outcomes 1 and 4, it is labelled *Number of obs.* In this example, it is 274;

ii. The point estimate and standard error for any HIVDR; here 0.642 and 0.0290, respectively;

iii. The point estimate and standard error for NNRTI HIVDR; here 0.511 and 0.0302, respectively.

d. To aggregate the data at a global level, it is also necessary to report the numerator of the prevalence estimate (and its associated standard error) and the denominator of the prevalence estimate (and its associated standard error). For Outcome 2a, the numerator is an estimate of the total number of treatment-naïve children less than 18 months of age newly diagnosed with HIV in the country during the 12-month survey period with known PMTCT exposure and any HIVDR (or NNRTI HIVDR for Outcome 2b). The denominator for Outcomes 2a/b is an estimate of the total number of treatment-naïve children less than 18 months of age newly diagnosed with HIV in the country during the 12-month survey period with known PMTCT exposure. The prevalence is equal to the numerator divided by the denominator. Select (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Totals*). In the Variables box in the Model tab, type or select `ANY_HIVDR_BN NNRTI_HIVDR_BN POP_SIZE`. In the *if/in/over* tab, type `PMTCT_EXPOSURE_RF == 1` (selects only children with known exposure) into the If: (*expression*) box. Press OK;

e. For Outcome 2a, the numerator estimate and its standard error are located in the row labelled `ANY_HIVDR_BN`; here, 1760 (standard error = 104.04). For Outcome 2b, the numerator estimate and its standard error are located in the row labelled `NNRTI_HIVDR_BN`; here, 1400 (standard error = 98.56). The denominator estimate and its standard error are located in the row labelled `POP_SIZE`; here, 2740 (standard error = 104.37).

```
. svy linearized, subpop(if PMTCT_EXPOSURE_RF == 1) : total ANY_HIVDR_BN NNR-
TI_HIVDR_BN POP_SIZE

Survey: Total estimation

Number of strata =       2          Number of obs    =      454
Number of PSUs   =     454          Population size  =     4540
                                    Subpop. no. obs  =      274
                                    Subpop. size     =     2740
                                    Design df        =      452

--------------------------------------------------------------
             |             Linearized
             |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
 ANY_HIVDR_BN |       1760   104.0375      1555.543    1964.457
NNRTI_HIVDR_BN |       1400   98.56066      1206.306    1593.694
      POP_SIZE |       2740    104.372      2534.886    2945.114
--------------------------------------------------------------
```

7.2 Reporting of HIVDR data

All countries are encouraged to report to WHO a dataset including: (1) individual patient information (demographic and matching laboratory data); (2) clinic data; and (3) survey variables discussed in Section 5.3, in addition to the patient sequences in FASTA file format. It is recommended that sequence IDs, case specimens IDs and specimen ID numbers be identical, as defined in Section 5.1.

In countries where individual patient information and sequences cannot be reported, survey outcomes and additional data on the prevalence of HIVDR in different subpopulations should be reported in an aggregated fashion. An Excel data collection and reporting tool will be available on the WHO website. Prevalence data should be accompanied by numerator, denominator, standard error of prevalence, standard error of numerator, and standard error of denominator, to allow pooling of regional and global data.

For this survey, the Stanford HIVdb[1] is used to classify HIVDR. This algorithm classifies HIVDR in five levels: susceptible, potential low-level, low-level, intermediate or high-level drug resistance. Sequences classified as susceptible and potential low-level resistance are considered to have "no HIVDR". The utilization of these different categories is summarized below.

HIVDR by individual drug

When reporting HIVDR by individual drug, sequences classified as low-level, intermediate or high-level resistance (according to the Stanford HIVdb) should be classified as "HIVDR". This classification applies to all drugs.

HIVDR by drug class

When reporting HIVDR by drug class, the following operational definitions for drug class should be used:

1. NNRTI class refers to any NVP or EFV;
2. Nucleoside reverse transcriptase inhibitor (NRTI) class refers to any NRTI;
3. Boosted PI class refers only to DRV/r, LPV/r or ATV/r;
4. Integrase inhibitor (INI) class refers to any INI.

Sequences classified as low-level, intermediate or high-level resistance (according to the Stanford HIVdb) should be aggregated as "HIVDR".

Any HIVDR

"Any HIVDR" is defined in sequences classified as low-level, intermediate or high-level resistance (according to the Stanford HIVdb) with respect to one or more of the following drugs: NVP, EFV, any N(t)RTI, DRV/r, LPV/r or ATV/r.[2]

In countries where these individual variables and sequences cannot be reported, aggregate data on HIVDR among treatment-naïve children less than 18 months newly diagnosed with HIV with known, no known, or unknown PMTCT exposure should be recorded and reported by drug. A standardized Excel-based reporting form will be available for download on the WHO HIVDR website at http://www.who.int/hiv/topics/drugresistance/en/index.html.

1. Available at: http://sierra2.stanford.edu/sierra/servlet/JSierra
2. INI should not be included.

8. STATISTICAL APPENDIX

8.1 Sample size calculations

To calculate the sample size, we simultaneously control the CI half-width of Outcomes 1a/b, 2a/b and 3a/b. All genotyped samples can contribute to Outcomes 1a/b (HIVDR regardless of PMTCT exposure), while only a subset of specimens can contribute to Outcomes 2a/b (HIVDR given known PMTCT exposure) and Outcomes 3a/b (HIVDR given no or unknown PMTCT exposure). The most precise outcomes will be 1a/b. The precision of Outcomes 2a/b and 3a/b will depend on which proportion of the population has known *versus* no or unknown exposure. We place a large upper bound on the CI half-width (11% or $L=0.11$), assuming that only 20% of samples had no or unknown exposure ($p_{no/\ unknown}=0.20$). If the balance is closer to 50%/50%, the CIs will be narrower. Similarly, the sample size is appropriate if 20% of samples had known PMTCT exposure (maximum imbalance 80%/20%).

We assume the prevalence of HIVDR is 50% ($P_{DR}=0.50$). This is the most conservative choice because it requires the largest sample size. No design effect is required because we are using a stratified design (population stratified on laboratory) in which the size of each population (number of case specimens per laboratory) is known prior to sampling. We further assume that 20% of genotypes will be unsuccessful ($p_{lab}=0.80$). The general form for the sample size is therefore:

$$N = \frac{3.84 \times p_{DR}\,(1 - p_{DR})}{L^2 \times p_{no/\ unknown} \times p_{lab}}$$

Entering in our assumed values, the required sample size is 496, rounded to 500.

8.2 Data analysis

8.2.1 Data analysis: introduction

We treat laboratories as a stratifying variable that divides the population of EID case specimens into well defined groups. We sample EID case specimens from each stratum (laboratory), and the number of case specimens collected from each laboratory is proportional to the total number of eligible case specimens available at that laboratory. We analyse the data using a survey framework accounting for stratification. If observations within strata are correlated, our analysis will be more efficient than a traditional analysis.

8.2.2 Data analysis: sampling weights for Outcomes 1, 2 and 3

M_i is a count of the number of EID case specimens available at laboratory i collected during a recent 12-month period. m_i is the number of case specimens successfully genotyped at laboratory . The sampling weight for a case specimen from laboratory i is $w_i = \frac{M_i}{m_i}$. Case specimens from the same laboratory are assigned the same weight.

8.2.3 Data analysis: Outcomes 1a and 1b

Outcome 1a is the overall prevalence of any HIVDR among all treatment-naïve children less than 18 months of age newly diagnosed with HIV, regardless of PMTCT exposure. Outcome 1b is the prevalence of HIVDR to NNRTI among all treatment-naïve children less than 18 months of age newly diagnosed with HIV, regardless of PMTCT exposure. These outcomes are analysed in a similar fashion. Data analysis for this and all additional outcomes can be conducted in Stata using the SVY utilities for a stratified random sample.[1] Even if Stata is not used to conduct the analysis, the Stata SVY manual section on variance estimation contains all necessary formulae for calculating the prevalence, variance and 95% CI of each outcome. The laboratory sampling weight is defined in Section 8.2.2.

---

1. StataCorp. 2013. Stata: Release 13. Statistical software. College Station, TX, StataCorp LP.

All genotyped specimens are defined as having either any HIVDR or NNRTI HIVDR, for Outcomes 1a and 1b, respectively (binary variable for HIVDR mutations = 1), or no detected HIVDR (binary variable for HIVDR mutations = 0). The prevalence is estimated using a ratio. The numerator is an estimate of the total number of EID case specimens in the country with HIVDR mutations during the survey period. The denominator is an estimate of the total number of EID case specimens in the country during the survey period. The variance is calculated using linearization. A 95% CI can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

8.2.4 Data analysis: Outcomes 2a and 2b

Outcome 2a is the prevalence of any HIVDR among treatment-naïve children less than 18 months of age newly diagnosed with HIV with known PMTCT exposure (maternal only, neonatal only or both). Outcome 2b is the prevalence of HIVDR to NNRTI among treatment-naïve children less than 18 months of age newly diagnosed with HIV with known PMTCT exposure (maternal only, neonatal only or both). These outcomes are analysed in a similar fashion. Data analysis is conducted using the same sampling weights and HIVDR outcomes as described for Outcomes 1a/b. The difference is that the population is restricted to case specimens with known PMTCT exposure (maternal only, neonatal only or both) using the subpopulation command in Stata. The prevalence is estimated using a ratio. The numerator is an estimate of the total number of case specimens in the country with known PMTCT exposure with HIVDR mutations during the survey period. The denominator is an estimate of the total number of case specimens in the country with known PMTCT exposure during the survey period.

8.2.5 Data analysis: Outcomes 3a and 3b

Outcome 3a is the prevalence of any HIVDR among treatment-naïve children less than 18 months of age newly diagnosed with HIV with no or unknown PMTCT exposure (maternal only, neonatal only or both). Outcome 3b is the prevalence of HIVDR to NNRTI among treatment-naïve children less than 18 months of age newly diagnosed with HIV with no or unknown PMTCT exposure (maternal only, neonatal only or both). These outcomes are analysed in a similar fashion. Data analysis is conducted using the same sampling weights and HIVDR outcomes as described for Outcomes 1a/b and 2a/b. The difference is that the population is restricted to case specimens with no or unknown PMTCT exposure using the subpopulation command in Stata. The prevalence is estimated using a ratio. The numerator is an estimate of the total number of case specimens in the country with no or unknown PMTCT exposure with HIVDR mutations during the survey period. The denominator is an estimate of the total number of case specimens in the country with no or unknown PMTCT exposure during the survey period.

8.2.6 Data analysis: sampling weights for Outcome 4

M_i is a count of the number of EID case specimens available at laboratory i collected during a recent 12-month period. n_i is the number of case specimens sampled at laboratory i with basic demographic information available. The sampling weight for a case specimen from laboratory i is $w_i = \frac{M_i}{n_i}$. Case specimens from the same laboratory are assigned the same weight.

8.2.7 Data analysis: Outcomes 4a, 4b and 4c

Outcomes 4a, 4b and 4c are the prevalence of "yes" PMTCT exposure, "no" PMTCT exposure and "unknown" PMTCT exposure to ARVs, respectively, among treatment-naïve children less than 18 months of age newly diagnosed with HIV. The sampling weight is defined in Section 8.2.6. All case specimens are defined as having known PMTCT exposure, no PMTCT exposure, or unknown PMTCT exposure (categorical variable). The prevalence of each category is estimated using a ratio. The variance is calculated using linearization. A 95% CI can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

For more information, contact:

World Health Organization
Department of HIV/AIDS
20, avenue Appia
1211 Geneva 27
Switzerland

E-mail: hiv-aids@who.int

www.who.int/hiv

ISBN 978-92-4-151254-1