World Health Organization

HIV DRUG RESISTANCE

CLINIC-BASED SURVEY OF ACQUIRED HIV DRUG RESISTANCE

AUGUST 2021

World Health Organization

HIV DRUG RESISTANCE

CLINIC-BASED SURVEY OF ACQUIRED HIV DRUG RESISTANCE

AUGUST 2021

Clinic-based survey of acquired HIV drug resistance

This publication is the update of the document published in 2014 entitled *Surveillance of HIV drug resistance in adults receiving ART (acquired HIV drug resistance)*

ISBN 978-92-4-003500-3 (electronic version)
ISBN 978-92-4-003501-0 (print version)

Suggested citation. Clinic-based survey of acquired HIV drug resistance. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

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

Layout by ACW, London

CONTENTS

ACKNOWLEDGEMENTS

Silvia Bertagnolio (World Health Organization (WHO) Department of Global HIV, Hepatitis and STI Programmes) conceived and coordinated the development of this document, in collaboration with Michael R. Jordan (Division of Geographic Medicine and Infectious Diseases, Tufts Medical Center, Boston, Massachusetts, United States of America, and the Levy Center for Integrated Management of Antimicrobial Resistance, Tufts University, Boston, Massachusetts, USA). WHO acknowledges the technical inputs of Bethany Hedt-Gauthier and Stephanie Wu (Harvard University, Cambridge, Massachusetts, USA), who developed the statistical methods. WHO acknowledges the technical inputs of Amalia Giron and Seth Inzaule (consultants, WHO Department of Global HIV, Hepatitis and STI Programmes). WHO thanks Juliana Da Silva, Sherry Pals, Elliot Raizes and Amitabh Suthar (United States Centers for Diseases Control and Prevention), Fatim Cham (WHO Regional Office for Africa) and Giovanni Ravasi (Pan American Health Organization) for their technical inputs to the initial draft of the document and the members of the WHO HIVResNet Technical Working Group on Surveillance and Monitoring – Chris Archibald, Santiago Avila, Emiliano Bisso, Mohammad Chakroun, Natalie Dean, David Dunn, Philippe Glaziou, Charles Holmes, Gillian Hunt, Derval Igoe, Mathieu Maheu-Giroux, Urvi Parikh, Andrew Phillips, Steve Reynolds – for their contributions.

ABBREVIATIONS AND ACRONYMS

ADR	acquired drug resistance
ART	antiretroviral therapy
ATV/r	atazanavir/ritonavir
CI	confidence interval
DBS	dried blood spot
DE	design effect
DRV/r	darunavir/ritonavir
DTG	dolutegravir
EFV	efavirenz
HIV	human immunodeficiency virus
HIVDR	HIV drug resistance
ICC	intracluster correlation
INI	integrase inhibitor
LPV/r	lopinavir/ritonavir
NRTI	nucleoside reverse-transcriptase inhibitor
NNRTI	non-nucleoside reverse-transcriptase inhibitor
NVP	nevirapine
PI	protease inhibitor
PI/r	ritonavir-boosted protease inhibitor
PPPS	probability proportional to proxy size
PR	protease
/r	with ritonavir boosting
RT	reverse transcriptase
SID	survey identification
VL	viral load
VNS	viral non-suppression

DEFINITIONS

Adult: generally, people 18 years of age and older; however, the definition may vary from country to country.

Children and adolescents: generally, individuals younger than 18 years of age; however, the definition may vary from country to country.

Viral load testing coverage: the proportion of all people receiving antiretroviral therapy who have at least one annual viral load test with classifiable results. Viral load testing coverage rates are usually derived from programmatic data.

Viral non-suppression: for the purposes of this survey, viral non-suppression is defined as HIV RNA ≥1000 copies/mL.

EXECUTIVE SUMMARY

The survey method outlined in this document uses a nationally representative design to estimate the prevalence of acquired HIV drug resistance among (1) adults and (2) children and adolescents receiving antiretroviral therapy (ART) with viral non-suppression (viral load ≥1000 copies/mL). This method is designed specifically for countries unable to implement WHO's 2021 laboratory-based survey of acquired HIV drug resistance using remnant viral load specimens because national viral load testing coverage is below the required threshold of 60%.

The survey outcomes are:

Outcome 1: Prevalence of viral suppression among adults and among children and adolescents receiving ART, regardless of the ART regimen.

Outcome 2: Prevalence of viral suppression among adults and among children and adolescents receiving a dolutegravir (DTG)-containing regimen.

Outcome 3: Prevalence of viral suppression among adults and among children and adolescents receiving ART, stratified by age band, gender, ART regimen (DTG-containing versus non-DTG-containing regimen), ART line, previous ART regimen, pregnancy status and breastfeeding status, if known.

Outcome 4: Prevalence of acquired HIV drug resistance among adults and among children and adolescents with viral non-suppression and receiving ART, regardless of the ART regimen.

Outcome 5: Prevalence of acquired HIV drug resistance to DTG among adults and among children and adolescents with viral non-suppression and receiving ART who are taking a DTG-containing regimen.

The survey method outlined in this document uses a two-stage cluster design. First, a set of ART clinics is randomly sampled, followed by sampling of individuals attending the selected clinics. ART clinics are sampled using probability proportional to proxy size (PPPS). Subsequently, consecutive individuals meeting the survey inclusion criteria are enrolled at sampled clinics and, as part of the survey, receive a viral load test and an HIV drug resistance test if their viral load is ≥1000 copies/mL. The method includes individuals on ART for any duration and stratifies by regimen to obtain overall and dolutegravir (DTG)-specific estimates of viral suppression and acquired HIV drug resistance. Sample sizes and statistical methods are designed to yield robust viral suppression estimates amongst people receiving DTG-containing and non-DTG-containing regimens. Finally, adults and children and adolescents are assessed separately in simultaneous surveys, as the prevalence of viral suppression and acquired HIV drug resistance, their determinants and public health actions may differ between these two groups.

Specimens are tested for HIV drug resistance to the HIV reverse transcriptase inhibitor (RT), protease inhibitor (PI) and integrase inhibitor (INI) drug classes at laboratories designated by WHO for the purpose of HIV drug resistance surveillance. De-identified survey participant-level variables, viral load and HIV drug resistance genotyping results are linked, and analyses are performed to obtain estimates for each outcome. Survey results are used to inform ART programme and public health policy with respect to optimal second-line ART regimens and benchmark national levels of HIV viral suppression amongst adults and children and adolescents receiving ART.

1. INTRODUCTION

As antiretroviral therapy (ART) for the treatment of HIV expands, it is essential to estimate, in a standardized and nationally representative manner, the extent to which acquired HIV drug resistance emerges in populations receiving therapy. Routine HIV drug resistance genotyping is neither recommended as part of the public health model of ART delivery, nor is it widely available in low- and middle-income countries.

The overarching purpose of this survey method is to generate a nationally representative prevalence estimate of acquired HIV drug resistance among ART-treated adults (and children and adolescents) in order to inform ART programme decision-making. In 2014 the World Health Organization (WHO) published a method for assessing acquired HIV drug resistance among people receiving ART by means of a nationally representative survey (*1*). This approach was designed to estimate viral suppression and acquired HIV drug resistance in populations on ART for 12 ± 3 months and 48 or more months.

In 2021, recognizing that an increasing number of people taking ART receive at least one viral load test per year, WHO published guidance on acquired HIV drug resistance surveillance that leverages remnant viral load specimens sampled from national viral load testing laboratories (*2*). However, in order to provide robust estimates of acquired HIV drug resistance using this method, viral load testing coverage of people receiving ART must equal or exceed 60%. Acknowledging that some countries have yet to attain this threshold, the method presented in this document provides an approach for countries with viral load testing coverage that is less than 60%.

The survey method outlined in this document uses a two-stage cluster design that is similar to the 2014 design (*1*). First, a set of ART clinics is randomly sampled, followed by sampling of individuals attending the selected clinics. As with the 2014 method, ART clinics are sampled using probability proportional to proxy size (PPPS). Subsequently, consecutive individuals meeting the survey inclusion criteria are enrolled at sampled clinics and, as part of the survey, receive a viral load test and an HIV drug resistance test if their viral load is ≥1000 copies/mL. In contrast to the 2014 method, this method includes individuals on ART for any duration and stratifies by regimen to obtain overall and dolutegravir (DTG)-specific estimates of viral suppression and acquired HIV drug resistance. In addition, sample sizes and statistical methods have been revised to yield robust viral suppression estimates for individuals receiving DTG-containing and non-DTG-containing ART. Finally, adults, on one hand, and children and adolescents, on the other, are assessed separately in simultaneous surveys, as the prevalence of viral suppression and acquired HIV drug resistance, their determinants and public health actions may differ between these two groups.

2. SURVEY METHODS

2.1 Overview

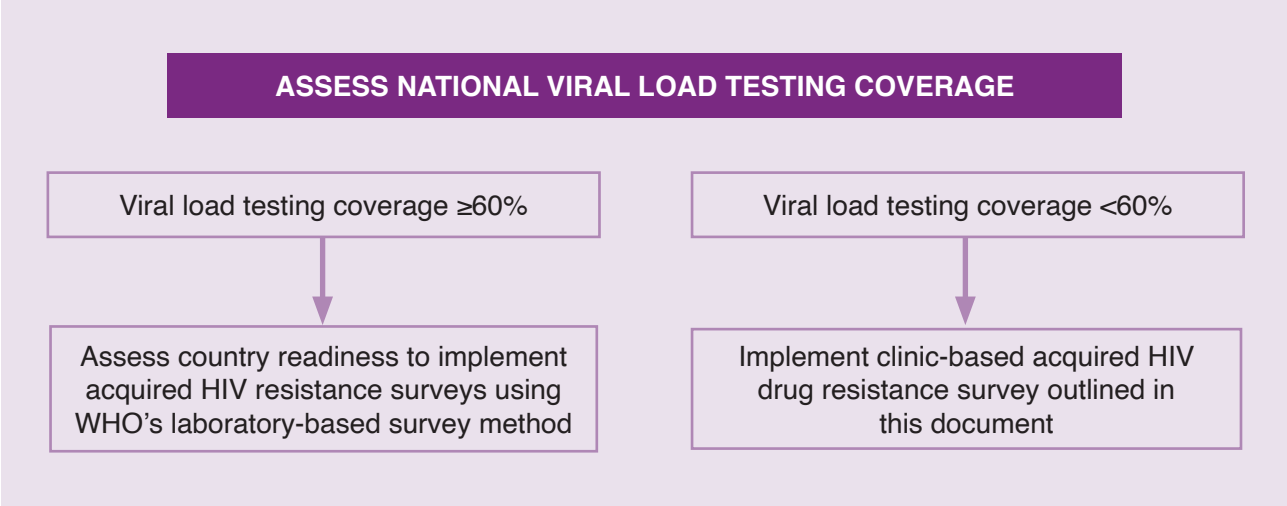
The survey is designed to precisely estimate the prevalence of HIV viral suppression among all people treated with ART, regardless of ART regimen, and specifically among those on DTG-containing regimens. The sampling design includes the DTG-containing versus non-DTG-containing regimens as stratifying variables. The approach described here can be used in both populations of adults and populations of children and adolescents. Section 2.6 provides specific guidance on the implementation of this survey in children and adolescents in a manner that minimizes the overall number of clinics to include in the combined samples of adults and children and adolescents.

Briefly, the survey uses a two-stage cluster design. First, ART clinics are randomly sampled for inclusion, using PPPS sampling. Second, eligible individuals within the clinic are sampled, with the first sample drawn from people receiving DTG-containing regimens ("DTG-eligible population") and the second sample drawn from people receiving non-DTG-containing regimens ("non-DTG eligible population"). The samples from the DTG-eligible population are used to estimate the prevalence of viral suppression among people taking DTG-containing regimens, and the combined samples (taken from DTG-eligible and non-DTG-eligible populations) are used to estimate the overall prevalence of viral suppression. Section 2.3 describes participant eligibility criteria.

Important points on implementing this survey method are summarized below:

- The methods described in this section are for countries with viral load testing coverage <60%. Countries with viral load testing coverage ≥60% are advised to conduct a readiness assessment to implement acquired HIV drug resistance surveys using remnant viral load specimens, following WHO recommendations (*2*) (Fig. 1).
- The survey method is a cross-sectional survey of individuals currently on ART, with eligible individuals prospectively enrolled.
- The recommended survey period is three months.
- The required patient-level survey variables are collected on a clinic-level data entry form at the time of survey enrolment and sample collection, with no patient-level identifying information being recorded for analysis. However, a link between the survey identification number and the ART number must be kept at the selected ART clinic to facilitate quality assurance of data and the return of HIVDR genotyping results, if desired.
- Specimen quality and quantity should be sufficient both for viral load testing and for possible HIVDR testing if an enrolled participant is found to have viral non-suppression. Dried blood spot (DBS) or plasma can be used as the specimen type for this survey. DBS has been shown to be a reliable specimen type for HIV drug resistance genotyping (*3*). DBS and plasma specimens should be collected and handled according to the 2020 WHO HIVResNet HIV drug resistance laboratory operational framework (*4*). Recognizing that viral load and drug resistance testing may often be performed in different laboratories, countries planning surveys of acquired HIV drug resistance are encouraged to collaborate with a WHO-designated laboratory in the survey planning stage to ensure that specimen collection, processing, handling, storage and shipment will yield quality HIV drug resistance testing results.
- Specimens from individuals without viral suppression (viral load ≥1000 copies/mL) should be tested in WHO-designated HIV drug resistance genotyping laboratories. These laboratories are members of the WHO HIVResNet Laboratory Network; they undergo a rigorous quality assurance process and participate in annual proficiency panel testing (4). Using WHO-designated laboratories guarantees quality-assured results for the purpose of public health surveillance. If a country does not have a WHO-designated laboratory for HIV drug resistance testing, it is encouraged to send specimens to a WHO-designated regional or specialized laboratory. A list of WHO-designated laboratories is available at the WHO HIV drug resistance website, https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/treatment/hiv-drug-resistance/laboratory-network.
- The reverse transcriptase, protease and integrase regions of the HIV-1 pol gene are sequenced using standard sequencing methods generating drug resistance information for HIV-1 nucleoside reverse-transcriptase inhibitors (NRTI), non-nucleoside reverse transcriptase inhibitors (NNRTI), ritonavir-boosted protease inhibitors (PI/r) and integrase inhibitors (INI).

Fig 1. Recommended acquired HIV drug resistance survey methods as based on national viral load testing coverage



2.2 Survey outcomes

Outcome 1: Prevalence of viral suppression among adults and among children and adolescents receiving ART, regardless of the ART regimen.

Outcome 2: Prevalence of viral suppression among adults and among children and adolescents receiving DTG-containing regimen.

Outcome 3: Prevalence of viral suppression among adults and among children and adolescents receiving ART, stratified by age band, gender, ART regimen (DTG-containing versus non-DTG-containing regimen), ART line, previous ART regimen, pregnancy status and breastfeeding status, if known.

Outcome 4: Prevalence of acquired HIV drug resistance among adults and among children and adolescents with viral non-suppression and receiving ART, regardless of the ART regimen.

Outcome 5: Prevalence of acquired HIV drug resistance to DTG among adults and among children and adolescents with viral non-suppression and receiving ART who are taking a DTG-containing regimen.

Outcomes 1 and 2 drive the overall design and sample size calculations.

2.3 Eligibility criteria

Individuals eligible for the survey must meet inclusion criteria and not meet exclusion criteria:

2.3.1 Inclusion criteria

- The person is receiving any ART regimen for the treatment of HIV and has been receiving it for at least three months.[1]
- The person provides informed consent (verbal or written).[2]

2.3.2 Exclusion criterion

- In countries where routinely used antibody tests differentiate between HIV-1 and HIV-2, individuals with HIV-2 or individuals with HIV-1/HIV-2 co-infection are excluded.

1 Individuals who have initiated ART and subsequently stopped it are eligible, providing that they restarted it at least three months prior to survey enrolment.
2 For surveys among children and adolescents, parental consent is required, and the participants' assent is required for children over the age of 7 years.

2.4 Survey sample size

2.4.1 Sample size design parameters

Table 1 summarizes the initial parameters for calculating the sample size. These assumptions determine the necessary sample sizes for viral suppression prevalence estimates in the overall sample and the DTG sample.

For both the overall and DTG-specific viral suppression prevalence estimates, the desired level of precision for the 95% confidence interval (CI) is ±5%. For overall viral suppression prevalence estimates, the expected prevalence varies depending on the context and regimens in use in the country. For this survey, the anticipated prevalence of viral suppression has been set at 85%. Designing the survey using an 85% expected prevalence ensures a confidence interval no wider than ±5% for any prevalence of 85% or higher, should other conditions around intracluster correlation be met (discussed further in section 2.4.3). For people receiving DTG-containing regimens, the anticipated prevalence of viral suppression is higher, set at 90%. Designing the survey using a 90% expected prevalence ensures a confidence interval no wider than ± 5% for any prevalence of 90% or higher, should other conditions around intracluster correlation be met (discussed further in section 2.4.3).

Table 1. Assumptions used to calculate sample size for viral suppression among adults or among adolescents and children

Assumptions	All individuals, regardless of ART regimen	Individuals receiving DTG-containing regimens
Expected prevalence of viral suppression	85%	90%
Desired absolute precision (95% CI half-width)	± 5%	± 5%

2.4.2 Allocating samples across strata

At each clinic, the sample size will be allocated by regimen (that is, the design assumptions separately determine how many people receiving DTG-containing regimens to sample and how many people receiving non-DTG-containing regimens to sample). The DTG-specific sample will be large enough to satisfy two requirements. First, the sample size must be large enough to satisfy viral suppression precision targets among adults and among children and adolescents taking DTG-containing regimens. Second, the sample size must be large enough that, when the DTG-specific sample is combined with the non-DTG sample, precision targets for the overall estimates of viral suppression are satisfied.

The sample size calculator first determines the per-clinic DTG-specific sample size to ensure that the first requirement is met (that is, that the sample size is large enough to satisfy the precision targets for DTG-containing regimens). However, if the proportion of adults or of children and adolescents taking DTG-containing regimens is small, the final clinic-level DTG-specific sample size may be inflated to ensure that the second requirement is met (that is, the sample size is large enough that, when combined with the non-DTG sample, the precision targets for the overall estimates are satisfied). Annex 2, section A2.1.4 provides details. The per-clinic sample sizes by regimen that satisfy both criteria are automatically calculated using the online WHO sample size calculator available at: https://worldhealthorg.shinyapps.io/ADR_ClinicBasedMethod/. Annex 3 provides an example of sample size calculations using the online WHO sample size calculator.

2.4.3 Adjusting sample sizes for clustering

Because ART clinics ("clusters") are sampled first, sample sizes must be adjusted for expected clustering in the viral suppression estimates. We make this adjustment by setting an anticipated level for the intracluster correlation coefficient (ICC). Two design options are available to countries; the difference between the two options is how much clustering is accommodated in the design through the value of the ICC. Option 1 accounts for higher levels of clustering; however, as a result, this option requires more clinics to be included as well as larger overall sample sizes. In contrast, Option 2 accounts for a lower level of clustering in the design and, as a result, requires fewer clinics and smaller sample sizes. The consequences of setting a lower value for the ICC are described more below under Option 2.

- **Option 1 (preferred):** In this first option, the ICC is set at 0.09, which is the median of the estimated country-level ICCs calculated from surveys of acquired HIV drug resistance implemented during the period 2016–2019

following WHO-recommended methods. Annex 2, section A2.1.2.1 provides details.

- **Option 2:** In this less desirable option, the ICC is set at 0.06, which is the 25th percentile of the estimated country-level ICCs from surveys of acquired HIV drug resistance implemented during the period 2016–2019 following WHO-recommended methods. Annex 2, section A2.1.2.2 provides details. In reality, the true ICC for the prevalence of viral suppression in a country is unknown; the anticipated ICC is a best assumption based on historical data. Setting the ICC to the 25th percentile provides design options that enhance feasibility for some countries. However, it is possible that an ICC of 0.06 is too low compared with the true ICC. In this case the resulting estimates and confidence intervals will still be correctly estimated; however, the confidence intervals will likely be wider than the targeted ± 5% precision.

2.4.4 Sample size inflation for viral load testing failure and imperfect weights

Since not all specimens will be successfully tested for viral load, the required sample sizes must be inflated to account for the laboratory testing failure rate. WHO recommends using an anticipated viral load testing failure rate of 10% in this calculation. In addition, because sampling is done proportionate to proxy size, rather than directly proportionate to size, updated weights must be included in the analysis. As a result, the variance of the final estimates will be increased. To accommodate this increased variance, the sample size is inflated by 1.5 (i.e., design effect (DE)=1.5). Annex 2, section A2.1.3 details both of these inflation factors. The resulting target sample sizes account for these inflation factors and are automatically reported in the online WHO sample size calculator (https://worldhealthorg.shinyapps.io/ADR_ClinicBasedMethod/).

2.4.5 Country-specific sample sizes

Country-specific sample sizes will vary, depending on (a) the number of clinics and the number of people receiving ART in the country; (b) the proportion of individuals on ART receiving DTG-containing regimens; and (c) the amount of clustering that the country can afford to accommodate in the survey design (as described in section 2.4.3). With the input of these parameters, the online WHO sample size calculator (https://worldhealthorg.shinyapps.io/ADR_ClinicBasedMehod/) automatically determines the minimum number of clinics that a country can select. The number of individuals per clinic and the total sample sizes will depend on the final number of clinics the country decides to sample. Annex 3 provides an example of sample size calculations using the online WHO sample size calculator.

Table 2 shows examples of two extreme scenarios: The first extreme is where 70% or more of the ART-treated population is receiving DTG-containing regimens, and the second extreme is where only 10% of the treated population is receiving DTG-containing regimens. For illustrative purposes, both scenarios assume an "infinitely" large number of clinics and people on ART. The sample sizes presented in Table 2 are generated under the first and preferred option for clustering, using the ICC=0.09. If 40 clinics are sampled, and if 70% of people are receiving DTG-containing regimens, then 28 individuals per clinic would need to be sampled (19 receiving DTG-containing regimens and nine receiving non-DTG-containing regimens), for a total sample size of 1120. If, instead, 10% of individuals are receiving DTG-containing regimens, then 37 per clinic would need to be sampled (12 receiving DTG-containing regimens and 25 receiving non-DTG-containing regimens), for a total sample size of 1480. In the bottom half of the table, sample sizes decrease when accounting for a smaller number of clinics (set at 300) and fewer people on ART (set at 20 000).

Table 2. Example of country-specific sample sizes under two extreme scenarios – (1) the proportion of the eligible population receiving DTG is equal to 70% and (2) the proportion of the eligible population receiving DTG is equal to 10%

Proportion of the eligible population receiving DTG is ≥70%					Proportion of the eligible population receiving DTG equals 10%				
"Infinitely" large number of clinics and people on ART									
	Sample size per clinic				Sample size per clinic				
Clinics	DTG	Non-DTG	Total	Total sample size	Clinics	DTG	Non-DTG	Total	Total sample size
40	19	9	28	1120	40	12	25	37	1480
45	14	6	20	900	45	9	18	27	1215
50	11	5	16	800	50	8	14	22	1100
55	9	4	13	715	55	7	11	18	990
If number of clinics=300 and the total number of people on ART=20 000									
	Sample size per clinic				Sample size per clinic				
Clinics	DTG	non-DTG	Total	Total sample size	Clinics	DTG	non-DTG	Total	Total sample size
40	12	5	17	680	40	5	15	20	800
45	10	4	14	630	45	4	12	16	720
50	8	4	12	600	50	4	10	14	700
55	7	3	10	550	55	4	9	13	715

DTG=dolutegravir

Table 2 is an example, starting at 40 sampled clinics. There is a minimum number of clinics that must be sampled. The online WHO sample size calculator will calculate this for countries. Countries then specify the number of clinics that they will sample, greater than or equal to this minimum number. Sampling more clinics is always preferable from a statistical perspective, particularly if there are anticipated large differences in viral suppression prevalence by clinic (and, thus, a large ICC). Increasing the number of clinics sampled will also decrease the total required sample sizes, as demonstrated in Table 2.

2.4.6 Precision of the acquired drug resistance estimates

The sample sizes for this activity are set to achieve precision of viral suppression estimates overall and within the population receiving DTG-containing regimens. However, other key outcomes are the acquired drug resistance (ADR) prevalence overall and the DTG-specific ADR prevalence in individuals receiving DTG-containing regimens (section 2.2, Outcomes 4 and 5). The online WHO sample size calculator reports the anticipated precision in these estimates, based on the following assumptions:

- 10% of samples will not have a viral load test result available for analysis (as described in section 2.4.4).
- Of those tested, the viral suppression rate is 85% overall and 90% in the population receiving DTG-containing regimens, consistent with the design parameters in Table 1.
- Of those with viral non-suppression (defined as VL ≥1000 copies/mL), the genotypic testing failure rate is 30%. This is consistent with the guidance for acquired HIV drug resistance surveillance leveraging remnant viral load specimens sampled from national viral load testing laboratories (2).
- The variance of the estimate is inflated by 1.5 due to cluster sampling, corresponding to a DE=1.5. There is currently no information on the ICC for ADR estimates; thus, this method uses a flat design effect adjustment.
- The expected prevalence of acquired HIV drug resistance among people with viral non-suppression is 50% overall, and the DTG-specific acquired drug resistance prevalence among those receiving DTG-containing regimens is 3.5%. This is consistent with the design parameters in the guidance for acquired HIV drug resistance surveillance leveraging remnant viral load specimens sampled from national viral load testing laboratories (*2*).

2.5 Sampling procedures

Clinics will be sampled using PPPS; namely, the probability of a clinic being sampled is the proportion of adults (or of children and adolescents) on ART at the clinic divided by the total number of adults (or of children and adolescents) on ART. Annex 1 provides more information on selecting the clinics to survey. The design assumes that the proportion of individuals receiving DTG- and non-DTG-containing regimens is constant across clinics, which is why this is considered probability proportional to the **proxy size**. The updated weights used in the analysis will reflect the fact that these proportions can, and likely do, vary by clinic.

Individuals within sampled clinics will be sampled using consecutive sampling, stopping when the DTG- and non-DTG target sample size per clinic (as determined by the online WHO sample size calculator) is reached. The actual size of the eligible population, by DTG- and non-DTG regimen, should be tracked for each sampled clinic over the three-month survey period to calculate the appropriate sampling weight for analysis. That is, all sampled clinics should continue to report the number of individuals receiving ART during the survey period, along with whether they are receiving a DTG- or a non-DTG-containing regimen, even after the target sample sizes are reached.

2.6 Survey of acquired HIV drug resistance among children and adolescents

Surveys of acquired HIV drug resistance among children and adolescents may be implemented separately from the adult survey or combined with the implementation of adult surveys. The operational differences are defined below.

Survey type: children and adolescents survey implemented separately

If the survey among children and adolescents is implemented separately from the survey among adults, all procedures for survey sample size, sampling and data analysis will be identical to those described for the adult survey, with the exceptions that the clinics included for sampling are clinics supporting children and adolescents (either combined with adults or child-only clinics) and survey participants are children and adolescents rather than adults.

Survey type: children and adolescents survey implemented in combination with an adult survey

Ideally, the same clinics sampled for the adult population could be used in the sample for the paediatric and adolescent population. This would be more affordable and easier to implement and supervise than sampling separate clinics. However, this is feasible only if children and adolescents are treated at the same clinics as adults and if the distribution of where children and adolescents are treated is similar to that of adults (for example, clinics where many adults are treated also have sizable paediatric and adolescent patient populations). Unfortunately, this is often not the case; thus, an alternative strategy has been devised where clinics that are sampled for the adult survey and serve both adults and children and adolescents will also be used for the survey among children and adolescents, with this set of clinics supplemented by additional clinics serving children and adolescents.

The strategy follows the processes outlined in sections 2.4.1–2.4.5 and is supported by the online WHO sample size calculator. The online WHO sample size calculator should be used twice: firstly, for the adult population and secondly for the child and adolescent population. This will result in two sets of sample sizes: (1) the number of clinics and the number of individuals per clinic (by regimen) for adults and (2) the number of clinics and the number of individuals per clinic (by regimen) for children and adolescents. The numbers are expected to be different between adults, on one hand, and children and adolescents, on the other, and presumably smaller for children and adolescents, as the overall number of children and adolescents on ART in a country is most likely smaller than the number of adults.

Next, countries should start by sampling adult clinics from the full set of clinics serving the adult population, using PPPS as described in section 2.5 and Annex 1. Likely, some of these sampled clinics also serve children and adolescents. These sampled clinics should be retained in the set of clinics included for the child and adolescent sample. The remaining number of clinics to be sampled for children and adolescents should be pulled from the full set of clinics serving children and adolescents, with those already included in the sample via the adult sampling process removed. Box 1 provides detailed instructions on implementation of the clinic sampling process for the survey among adults and the survey among children and adolescents. Fig. 2 presents a visual representation of the clinic sampling process described in Box 1.

There are two additional considerations when implementing this process, reflecting two scenarios. The first scenario is where the number of clinics sampled in the adult sampling process that also provide ART to children and adolescents is greater than the number of clinics recommended for children and adolescents. In this case it is recommended to supplement the sample with at least five additional clinics, sampled from all clinics serving children and adolescents (either combined with adults or child and adolescent-only clinics), excluding the clinics serving children and adolescents selected via the adult sampling process. This will ensure that clinics that are serving only children and adolescents have some probability of being sampled, thus yielding an unbiased ADR estimate.

A second scenario is that some of the clinics serving both adults and children and adolescents will, in reality, provide ART to very few children and adolescents. For the purposes of feasibility, it is recommended to classify these clinics as "adult-only clinics", and they may be excluded from the child and adolescent sampling frame following a process analogous to sampling small clinics, as discussed in section 2.7.

Box 1. How to calculate overall and clinic-specific sample sizes for surveys of acquired HIV drug resistance in adults and in children and adolescents and how to sample ART clinics for the survey

1) Calculate sample sizes for the survey among adults and for the survey among children and adolescents using the online WHO sample size calculator:

a) Adult sample sizes:

- Inputs: number of adults on ART, number of clinics serving adults and proportion of adults on ART who are taking DTG-containing regimens.
- Outputs: minimum number of clinics serving adults to be sampled, and sample sizes per clinic based on the final determined number of clinics serving adults to be sampled.

b) Child and adolescent sample sizes:

- Inputs: number of children and adolescents on ART, number of clinics serving children and adolescents and proportion of children and adolescents on ART who are taking DTG-containing regimens.
- Outputs: minimum number of clinics serving children and adolescents to be sampled, and sample sizes per clinic based on the final determined number of clinics serving children and adolescents to be sampled.

2) Sample clinics for survey in adults:

a) Construct adult sampling frame of clinics serving adults (adults only and adults + children and adolescents) using the number of adults on ART by clinic in prior year.

b) Sample clinics using PPPS from the adult sampling frame (Annex 1).

c) Keep track of the sampled clinics that serve children and adolescents as well as adults; they will be carried forward to the child and adolescent survey.

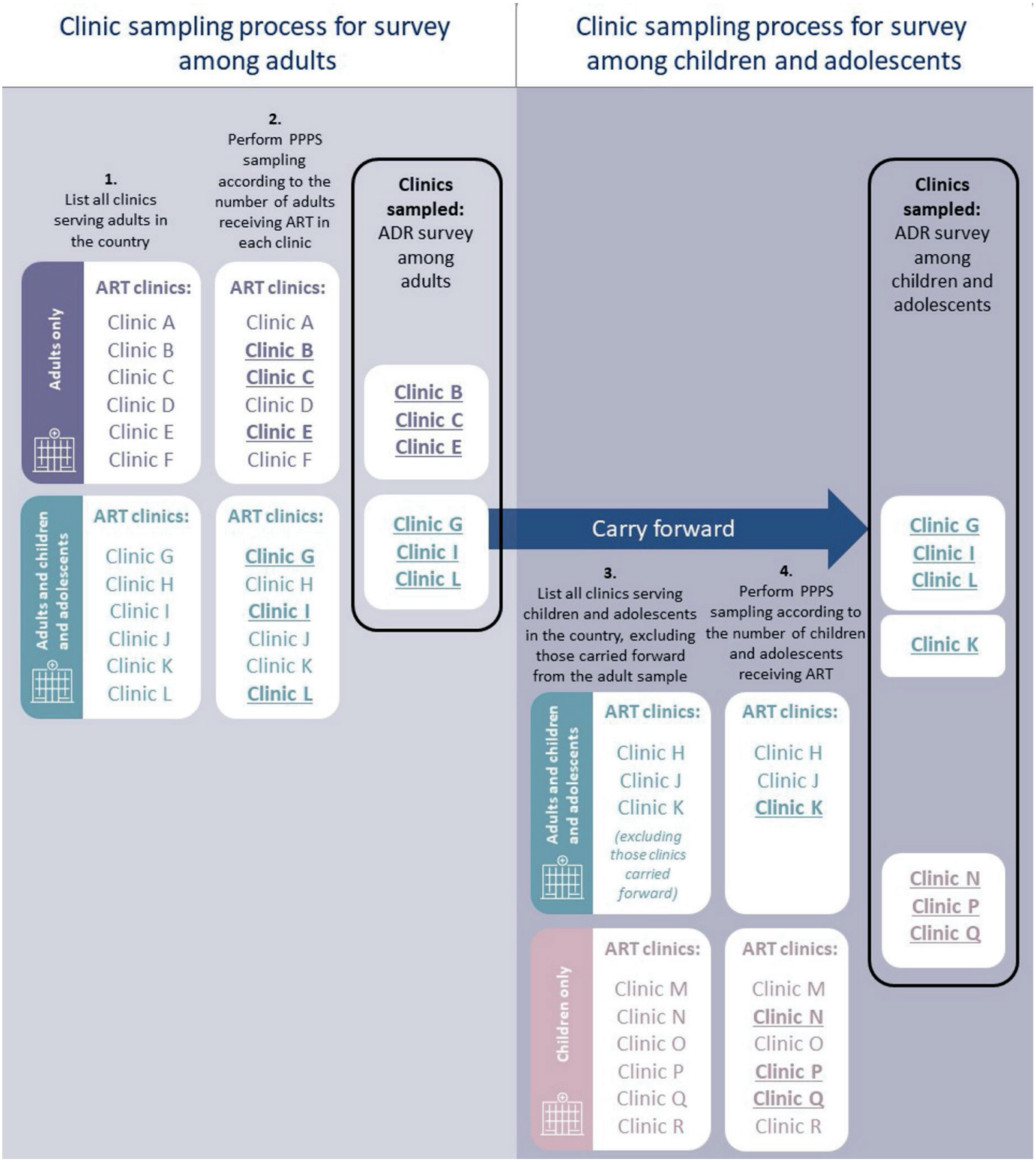
3) Sample clinics for survey in children and adolescents:

a) Carry forward clinics serving adults and children and adolescents that were sampled in the adult survey. These clinics are automatically included in the child and adolescent survey and, therefore, are excluded from the child and adolescent sampling frame.

b) Construct a child and adolescent sampling frame of clinics serving children and adolescents (children and adolescents only and adults + children and adolescents), excluding those carried forward from the adult sampling frame.

c) Sample the remaining clinics (the number recommended in Step 1b or the actual number of clinics serving children and adolescents to be sampled, if different, minus those carried forward in Step 2c) from the child and adolescent sampling frame using PPPS based on the number of children and adolescents on ART at the clinic in the prior year (Annex 1).

Fig. 2. Clinic sampling process for survey among adults and survey among children and adolescents



2.7 Sampling small clinics

Some countries may have a number of clinics with extremely small populations of people receiving ART or clinics that may be difficult to access for a variety of reasons, including political instability or geographical remoteness. Although not advisable, such countries may consider excluding some of these clinics from the systematic sampling table due to logistical and under-enrolment issues. In general, if these clinics represent less than 10% of the population receiving ART in the country, countries may choose to exclude these clinics from the systematic sampling table. This 10% threshold seeks to limit the potential bias that such exclusion may introduce in the final results. In this case exclusion from the systematic sampling table should be done *a priori* (and not after the clinic has been sampled). A list of all excluded clinics and reasons for their exclusion should be reported in any resulting technical report. In contrast, if more than 10% of the population of interest attends these clinics, it is not advisable to exclude these clinics from the pool of clinics that can be sampled for the survey. In general, if the excluded individuals have a different prevalence of viral suppression than the observed individuals, the national prevalence estimate will be biased. This should be taken into consideration when interpreting these results.

If the required sample size per clinic is not achieved during the maximum survey enrolment period, the survey will not achieve the predetermined sample size. If the amount of under-enrolment is minimal, under-enrolment will not greatly affect the precision of the survey. However, if under-enrolment is large, the resulting prevalence estimate of viral load suppression will have a wider confidence interval than originally planned. Before starting the survey, countries should assess whether sampled clinics are expected to be able to enrol the required number of people during the three-month survey period. The question to be asked at the national level is: How likely is it that the required number of people receiving ART (DTG- and non-DTG-containing) will be observed at each selected clinic during the survey period? When assessing the enrolment capabilities of the selected clinics, countries can use the number of people on ART at the end of a previous one-year period and divide by four, assuming a uniform distribution of patient visits during the year and given that the survey is expected to last three months.

If a country expects to encounter significant under-enrolment (for example, because relatively small clinics were selected during random sampling), the expected difference should be equally distributed to larger clinics. For example, if the expected under-enrolment is 40 people, and there are five large clinics, an additional eight individuals should be sampled from each large clinic. The proposed strategy to create sampling weights will no longer be exactly inverse probability weights, but it will be sufficient to approximate the overall sampling probability, even with the excess sample reassigned.

2.8 Regional representation

This method generates nationally representative estimates. Countries wishing to develop region-specific estimates of acquired HIV drug resistance and viral load suppression should implement a separate survey in each area of interest.

2.9 Observed sample size

Not all sampled specimens will successfully be tested for viral load. The observed sample size is the number of people sampled who are successfully tested and for whom viral load test results are available for analysis. All of these samples are included in the analysis.

Ideally, the observed sample size is at least as large, if not larger, than the sample size required to meet the constraints. If the observed sample size is smaller than the required sample size, because the viral load testing failure rate exceeds 10%, then the confidence intervals may be wider than specified in the design. The analysis below remains statistically correct, however, whether or not the observed sample size is larger or smaller than the required sample size.

2.10 Data analysis

The statistical analysis for the viral suppression outcomes will account for clustering by the selection of clinics, stratification by individuals receiving DTG-containing or non-DTG-containing regimens, the sampling weights reflecting the true sampling probability (correcting for the fact that PPPS sampling was used instead of probability proportional to size (PPS)), nonresponse due to laboratory failure, and the fact that samples are drawn from finite populations. The statistical analysis for the ADR outcomes will use a partially weighted analysis to account for stratification by individuals receiving DTG-containing or non-DTG-containing regimens and nonresponse due to genotyping failure. Full inverse probability sampling weights will not be used for the ADR outcomes because the small sample sizes would likely result in unstable estimates with high variability. Annex 2, sections A2.3-A2.4 provide full formulas and Stata code that automate these calculations. Prevalence estimates and corresponding 95% confidence intervals are calculated.

3. IMPLEMENTATION CONSIDERATIONS

3.1 Duration of the survey, patient screening and sampling

To ensure that results are available to decision-makers in a timely fashion, the duration of patient sampling is limited to a maximum of three months. Once ART clinics to be included in the survey have been selected, a convenient starting date for the survey is chosen. All individuals attending the sampled clinics should be screened using the inclusion and exclusion criteria outlined in sections 2.3.1 and 2.3.2, respectively. All eligible individuals should be consecutively enrolled until the country-specific sample size per clinic is reached or until the maximum enrolment period of three months has passed. Individual clinics can stop sampling if the desired sample size is reached earlier; however, **clinics need to continue to count the number of eligible individuals, by regimen, that are seen during the survey period.**

3.2 List of variables to be collected

This section describes the minimal set of information that must be captured in the survey database.

Survey type: adults

3.2.1 Required patient-level information

a. Clinic ID

b. Participant survey ID (see Box 2 for identification conventions)

c. Date of birth or age

d. Date when ART was first initiated

e. Current ART regimen: list the names of each ARV drug

f. Current ART line (first-line/second-line/third-line/unknown)

g. Gender (female, male, other)

h. Viral load (VL) testing successful and results available? (IRQ=In range; BLQ=Below lower limit of quantitation; AUQ=Above upper limit of quantitation; UNK=Result not available)

i. VL result (copies/mL) from survey blood draw (VL is entered as copies/mL if the result is within assay quantitation limits. If the result is below the lower limit of quantitation, the lower limit is entered. If result is above the upper limit of quantitation, the upper limit is entered.)

j. If VL ≥1000 copies/mL, were reverse transcriptase and protease (RT/PR) regions of HIV-1 pol gene successfully sequenced? (Indicates whether the viral genotype was successfully sequenced, and data are available. Possible values: SUC=Sequencing successful; UNS=Sequencing unsuccessful; SNA=Sequencing not attempted; UNK=Sequencing status unknown).

k. If VL ≥1000 copies/mL, was the INI region of pol gene successfully sequenced? (Indicates whether the viral genotype was successfully sequenced, and data are available. Possible values: SUC=Sequencing successful; UNS=Sequencing unsuccessful; SNA=Sequencing not attempted; UNK=Sequencing status unknown).

3.2.2 Optional patient-level information

a. Prior ART regimen: list the name of each ARV drug

b. Breastfeeding status (patient is currently breastfeeding: Y=Yes/N=No/UNK=Unknown)

c. Pregnancy status (patient is currently pregnant: Y=Yes/N=No/UNK=Unknown).

3.2.3 Clinic-level information

a) Clinic ID

b) Total number of adults taking DTG-containing regimens, observed during the three-month survey period

c) Total number of adults taking non-DTG-containing regimens, observed during the survey period[1]

d) Clinic size as stated in the table used for systematic sampling (Annex 1 presents an example of a systematic sampling table.)

3.2.4 Country-level information used to inform sample size calculations at time of survey design

a. Survey type (that is, ADR)

b. Total number of clinics serving adults in the country

c. Total number of adults receiving ART in the country

d. Total number of clinics serving adults in the sampling table (Annex 1 presents an example of a systematic sampling table.)

e. Total number of adults receiving ART as defined in the sampling table

f. Total number of clinics serving adults to be sampled (may be greater than the total number of clinics participating in the survey, as some clinics may be sampled more than once)

g. Sampling interval from systematic sampling table for adult survey

h. Proportion of adults receiving ART who are on DTG-containing regimens in the entire country.

1 There are two ways to ascertain the total number of adults (or children and adolescents) taking DTG- and non-DTG-containing regimens who are observed during the survey period. Countries may choose to keep a screening log of all eligible individuals presenting to the clinic on or after the survey start date. The screening log would capture the relevant population (adult or else child and adolescent) and enumerate them by regimen type (that is, DTG-containing ART versus non-DTG-containing ART). Alternatively, countries may choose to query available clinic-level or pharmacy programmatic data documenting the total number of adults (and of children and adolescents) receiving DTG- and non-DTG-containing regimens over the last 12 months and divide by four, yielding an estimated number observed during the prospective survey period.

Survey type: children and adolescents, separate implementation

If the survey of children and adolescents is implemented separately from the survey of adults, rather than combined with it, the patient-, clinic-, and national-level variables necessary are analogous to those listed for the adult survey and described in sections 3.2.1–3.2.4, except that they will be the corresponding child and adolescent variables. For example, instead of "total number of clinics serving adults in the country," the survey will collect "total number of clinics serving children and adolescents in the country."

Survey type: children and adolescents, combined implementation

If the survey of children and adolescents is implemented alongside the survey of adults, as detailed in section 2.6, the information listed below, in sections 3.2.5–3.2.8, must be captured.

3.2.5 Required patient-level information

a. Clinic ID
b. Participant survey ID (see Box 2 for identification conventions)
c. Date of birth or age
d. Date when ART was first initiated
e. Current ART regimen: list the name of each ARV drug
f. Current ART line (first-line/second-line/third-line/unknown)
g. Gender (female, male, other)
h. VL testing successful and results available? (IRQ=In range; BLQ=Below lower limit of quantitation; AUQ=Above upper limit of quantitation; UNK=Result not available)
i. VL result (copies/mL) from survey blood draw (VL is entered as copies/mL if the result is within assay quantitation limits. If result is below the lower limit of quantitation, the lower limit is entered. If result is above the upper limit of quantitation, the upper limit is entered.)
j. If VL ≥1000 copies/mL, were reverse transcriptase and protease (RT/PR) regions of HIV-1 pol gene successfully sequenced? (Indicates whether the viral genotype was successfully sequenced, and data are available. Possible values: SUC=Sequencing successful; UNS=Sequencing unsuccessful; SNA=Sequencing not attempted; UNK=Unknown status of sequencing).
k. If VL ≥1000 copies/mL, was the INI region of pol gene successfully sequenced? (Indicates whether the viral genotype was successfully sequenced, and data are available. Possible values: SUC=Sequencing successful; UNS=Sequencing unsuccessful; SNA=Sequencing not attempted; UNK=Sequencing status unknown).

3.2.6 Optional patient-level information

a. Prior ART regimen: list the name of each ARV drug
b. Breastfeeding status (patient is currently breastfeeding: Y=Yes/N=No/UNK=Unknown)
c. Pregnancy status (patient is currently pregnant: Y=Yes/N=No/UNK=Unknown

3.2.7 Clinic-level information

a. Clinic ID
b. Total number of children and adolescents taking DTG-containing regimens observed during the survey period
c. Total number of children and adolescents taking non-DTG-containing regimens, observed during the survey period[1]
d. Adult clinic size (number of adults on ART as used in the systematic sampling table for the adult survey). This is the historical clinic size used when sampling clinics and not the number on ART during the survey period.
e. Child and adolescent clinic size (number of children and adolescents on ART as used in the systematic sampling table for the child and adolescent survey). This is the historical clinic size used when sampling clinics and not the number on ART during the survey period.
f. Clinic was selected through the adult sampling frame (Y=Yes/N=No)
g. Clinic serves both adults and also children and adolescents (Y=Yes/N=No)

3.2.8 Country-level information used to inform sample size calculations at time of survey design

a. Survey type (that is, ADR)
b. Total number of clinics serving adults in the country
c. Total number of adults receiving ART in the country
d. Total number of clinics serving adults in the adult sampling table (Annex 1 presents an example of a systematic sampling table.)
e. Total number of adults receiving ART as defined in the adult sampling table
f. Total number of clinics serving adults to be sampled (may be greater than the total number of clinics participating in the survey, as some clinics may be sampled more than once)
g. Sampling interval from systematic sampling table for adult survey
h. Proportion of adults receiving ART who are on DTG-containing regimens in the entire country
i. Total number of clinics serving children and adolescents in the country
j. Total number of children and adolescents receiving ART in the country
k. Total number of clinics serving children and adolescents in the child/adolescent sampling table (excluding clinics selected for the adult survey)

1 There are two ways to ascertain the total number of adults (or children and adolescents) taking DTG- and non-DTG-containing regimens who are observed during the survey period. Countries may choose to keep a screening log of all eligible individuals presenting to the clinic on or after the survey start date. The screening log would capture the relevant population (adult or else child and adolescent) and enumerate them by regimen type (that is, DTG-containing ART versus non-DTG-containing ART). Alternatively, countries may choose to query available clinic-level or pharmacy programmatic data documenting the total number of adults (and of children and adolescents) receiving DTG- and non-DTG-containing regimens over the last 12 months and divide by four, yielding an estimated number observed during the prospective survey period.

l. Total number of children and adolescents receiving ART in the children and adolescent sampling table (excluding children and adolescents receiving ART in clinics selected for the adult survey)

m. Total number of clinics serving children and adolescents to be sampled (may be greater than the total number of clinics participating in the survey, as some clinics may be sampled more than once)

n. Sampling interval from the children and adolescent sampling table

o. Proportion of children and adolescents receiving ART who are on DTG-containing regimens in the entire country

Box 2. Assigning a unique participant survey identification (SID) number, or unique survey ID

This box summarizes the convention for assigning unique SIDs to participants enrolled in this survey. The SID is 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 ("-"): country abbreviation: the International Organization for Standardization's standard 3-letter abbreviation - survey type: ADR - year survey started - site abbreviation (a 3-letter abbreviation for the site, unique within the country; by default, the first three letters of the site name unless this is not unique) - 4-digit unique patient number, that is, a consecutive unique patient number assigned to a participant at that site. A lower case "-a" denotes adults and a lower case "-c" denotes children and adolescents. For example, if the "University HIV Clinic" is a site participating in a national survey of ADR in adults in South Africa in 2021, a participant's ADR-SID would look like this: ZAF-ADR-2021-UHC-0001-a. Using this unique survey identifier is required if using the WHO HIV drug resistance database, which supports cleaning and quality assurance of both de-identified patient data and HIV sequences, thus enabling data analysis and quality-assured results.

3.3 Repeating the survey

This survey is designed to make possible the assessment of trends in the prevalence of viral load suppression and acquired HIVDR in populations receiving ART. Thus, it should be repeated periodically, generally every three years or more often. Countries implementing this methodology do so because viral load testing coverage is <60%. If a country implements this method, it is advised to strengthen, in parallel, its viral load testing coverage and to reassess its ability to implement WHO's laboratory-based survey of acquired HIV drug resistance using remnant viral load specimens (*2*) at the time of the next round. If, at a future date, a country achieves viral load testing coverage of ≥60%, it should implement the WHO's laboratory-based survey of acquired HIV drug resistance using remnant viral load specimens (*2*). If, however, the country continues to have viral load testing coverage of <60%, it should repeat this method, and, when doing so, it must update the sampling table and perform a new random sample of clinics to ensure that the new survey adequately reflects changes in the ART programme. Annex 4 presents an example of a budget.

3.4 Data extraction

The same unique survey identification number described in Box 2 must be assigned to the de-identified survey participant-level data, the participant's blood specimen and subsequent FASTA file header of the drug resistance genotype. All participant-, clinic- and national-level data should be entered into a spreadsheet-based data capture tool developed by WHO for this purpose (https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/treatment/hiv-drug-resistance/hiv-drug-resistance-surveillance/surveillance-of-acquired-hiv-drug-resistance-in-populations-receiving-art).

3.5 Reporting data

All countries are encouraged to report to WHO a dataset consisting of (1) individual de-identified participant-level survey information, (2) viral load laboratory-level data and (3) HIV sequences in FASTA file format. It is recommended that data identifiers follow the WHO convention as defined in Box 2. An Excel data upload template is available for download from the WHO HIV drug resistance database (https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/treatment/hiv-drug-resistance/hiv-drug-resistance-surveillance).

When reporting resistance to a specified antiretroviral drug, sequences classified as having predicted low-level, intermediate or high-level resistance (according to the Stanford HIVdb (*5*)) should be classified as drug resistant. Equally, this classification applies when reporting resistance to all drug classes.

When reporting HIV drug resistance by drug class, the following operational definitions for drug class are used:

- Any HIV drug resistance is defined as resistance to nevirapine (NVP), efavirenz (EFV), any NRTI, darunavir/ritonavir (DRV/r), lopinavir/ritonavir (LPV/r), atazanavir/ritonavir (ATV/r) or any INI.
- Resistance to the NNRTI drug class is defined as resistance to NVP or EFV.
- Resistance to the NRTI class is defined as resistance to any NRTI.
- Resistance to the ritonavir-boosted protease inhibitor drug class is defined as resistance to DRV/r, LPV/r or ATV/r.
- Resistance to the INI drug class is defined as resistance to any INI.

3.6 Implementation overview: practical guide

This section provides a practical overview to implementation. In this example a national HIV drug resistance working group in a country meets on 1 January 2021 and plans to implement an acquired HIV drug resistance survey. Funding is available until December 2021. The working group follows the steps below:

1. The working group first determines the national viral load testing coverage. If it is ≥60%, then the country uses WHO's laboratory-based acquired HIV drug resistance survey, which leverages remnant viral load specimens (2). If viral load testing coverage is <60%, then the country proceeds with this method.

2. Using the online WHO sample size calculator, the working group estimates, based on country data, the sample sizes, and the minimum number of clinics to be sampled both for adults and for children and adolescents and the budget for the survey. (Annex 4 provides an example of a budget.) **Note that two sampling frames – a sample of clinics treating adults and a sample of clinics treating children and adolescents – will be required if implementing the survey in children and adolescents.** While countries may choose to implement the survey in more than the minimum number of clinics, it is not statistically prudent to implement it in fewer.

3. Decide whether the available budget permits the implementation of a simultaneous survey among children and adolescents.

4. Selection of the clinics to be sampled proceeds by PPPS sampling, as detailed in Annex 1.

5. The survey period is three months. The working group selects the study start date. As per section 3.1, all individuals coming to the selected clinics on or after the survey start date are screened for eligibility and are asked for informed consent (oral or written).

6. All eligible individuals are consecutively enrolled until the required sample size for each population (DTG and non-DTG) per clinic is reached or until the 3-month enrolment period has passed. Individual clinics can stop sampling if the desired sample size is reached earlier. However, per sections 3.2.2 and 3.2.5 (children and adolescent), the total number of individuals receiving DTG- and non-DTG-containing ART who attend the clinic during the three-month survey period must be recorded, as this information is required in the analysis phase.

7. For eligible individuals, the survey should proceed in two steps:

   Step 1: Consent and required survey participant-level information obtained (section 3.2)

   Step 2: Blood drawn for viral load and HIV drug resistance genotyping

8. Once specimen collection is completed, specimens with VL ≥1000 copies/mL are sent to a WHO-designated laboratory for HIV drug resistance genotyping.

9. The WHO-designated laboratory performs HIV drug resistance genotyping, and de-identified patient information is entered into upload templates provided by WHO (section 3.5) for use with the WHO HIV drug resistance database.

10. De-identified patient data and HIV drug resistance genotypes are uploaded into the WHO HIV drug resistance database for cleaning and quality assurance of both de-identified demographic information and sequences.

11. The data are analysed.

12. The working group writes a national report, and data are disseminated and used for public health and ART programme policy-making.

4. REFERENCES

1. Surveillance of HIV drug resistance in adults receiving ART (acquired HIV drug resistance). Geneva: World Health Organization; 2014.

2. Laboratory-based survey of acquired HIV drug resistance using remnant viral load specimens. Geneva: World Health Organization; 2021.

3. Bertagnolio S, Parkin NT, Jordan M, Brooks J, Garcia-Lerma JG. Dried blood spots for HIV-1 drug resistance and viral load testing: a review of current knowledge and WHO efforts for global HIV drug resistance surveillance. AIDS Rev. 2010;12(4):195-208.

4. WHO HIVResNet HIV drug resistance laboratory operational framework. Second ed. Geneva: World Health Organization; 2020. p. 82.

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

ANNEX 1. SELECTING THE CLINICS TO SURVEY

A1.1 Systematic sampling of clinics

This section describes how to sample clinics from the list of all ART clinics in the country. Sampling of clinics is performed using PPPS to generate a random sample of clinics.

To execute systematic sampling, all clinics providing ART in the country are listed (Table A1.1). Operationally, a) list all eligible clinics providing ART along with the number of people (adults or children and adolescents) on ART at the end of the previous calendar year at each clinic (to reflect the relative sizes of the patient populations), b) calculate the cumulative population size for each clinic listed (described below), c) determine the sampling interval, d) pick a random starting-point, and e) select clinics based on the random starting-point, sampling interval and cumulative population size. Table A1.1 below illustrates these steps in greater detail.

Detailed instructions:

1. Generate a facilities list of all ART clinics in the country.

2. Record the number of people receiving ART at the end of the previous calendar year, by clinic. Starting at the top of the table, calculate the cumulative eligible population size for each clinic in another column. The cumulative eligible population size is the size of the clinic plus the size of all clinics previously listed in the table.

3. Determine the sampling interval by dividing the cumulative population size of all listed clinics by the number of clinics to be sampled. In the case of our example, the cumulative population size is 13 666 and the number of clinics to be sampled is 20. Therefore, the sampling interval is 13 666/20=683.3, rounded to 683.

4. Pick a random starting point. To select the first clinic, obtain a random number between 1 and the sampling interval 683. A random number generator can be found at http://www.random.org/. For illustration, the random number obtained in this example was 500.

5. Select clinics based on the random starting point, sampling interval and cumulative population size.

   a. Select the first clinic in which the size of the cumulative number treated is greater than or equal to the random number. The cumulative population size passes 500 in Clinic E. Because Clinic E is the first clinic in which the cumulative size is greater than or equal to the random start, Clinic E is selected.

   b. Add the initial random number and the sampling interval (500+683=1183), and then select the first clinic listed in which the cumulative total is greater than or equal to this number (1183). The cumulative size for Clinic F is 856, which is less than 1183 and so Clinic F is not sampled. The cumulative size for Clinic G is 1209, making clinic G the first clinic with cumulative size greater than or equal to 1183. Thus, Clinic G is selected. Continue adding the sampling interval to the result obtained until all 20 clinics have been selected.

It is possible for a clinic to be selected more than once if its eligible population size is larger than the sampling interval. In the example in Table A1.1, Clinic S is selected twice. If a clinic is picked twice, then twice the sample size must be taken from this clinic. For example, if the sample size is 23 people per clinic, then the sample size for that clinic is 46.
If a clinic is picked k times, then k times the sample size must be taken. The result is that fewer than 20 unique clinics are sampled. In our example, 19 clinics are sampled.

Table A1.1. Systematic sampling table for clinic selection

Clinic name	Number of people on ART at the end of previous calendar year	Cumulative total of eligible individuals	Selection	Sample clinic
Clinic A	300	300		
Clinic B	111	411		
Clinic C	53	464		
Clinic D	20	484		
Clinic E	16	500	500* (random start)	Clinic 1
Clinic F	356	856		
Clinic G	353	1209	500+683**=1183	Clinic 2
Clinic H	125	1334		
Clinic I	45	1379		
Clinic J	604	1983	1183+683=1866	Clinic 3
Clinic K	600	2583	1866+683=2549	Clinic 4
Clinic L	400	2983		
Clinic M	383	3366	2549+683=3232	Clinic 5
Clinic N	201	3567		
Clinic O	115	3682		
Clinic P	105	3787		
Clinic Q	99	3886		
Clinic R	25	3911		
Clinic S	687	4598	3232+683=3915	Clinic 6 (selected twice)
			3915+683=4598	
Clinic T	633	5231		
Clinic U	585	5816	4598+683=5281	Clinic 7
Clinic V	651	6467	5281+683=5964	Clinic 8
Clinic W	517	6984	5964+683=6647	Clinic 9
Clinic X	353	7337	6647+683=7330	Clinic 10
Clinic Y	330	7667		
Clinic Z	279	7946		
Clinic AA	167	8113	7330+683=8013	Clinic 11
Clinic BB	630	8743	8013+683=8696	Clinic 12
Clinic CC	464	9207		
Clinic DD	158	9365		
Clinic EE	33	9398	8696+683=9379	Clinic 13
Clinic FF	688	10086		
Clinic GG	598	10684	9379+683=10062	Clinic 14
Clinic HH	556	11240	10062+683=10745	Clinic 15
Clinic II	465	11705	10745+683=11428	Clinic 16
Clinic JJ	399	12104		
Clinic KK	285	12389	11428+683=12111	Clinic 17
Clinic LL	181	12570		
Clinic MM	143	12713		
Clinic NN	668	13381	12111+683=12794	Clinic 18
Clinic OO	285	13666	12794+683=13477	Clinic 19

*500=random start; generated by www.random.org
**683=sampling interval

ANNEX 2. STATISTICAL METHODS

This annex provides the statistical details of the clinic-based survey approach.

Table A2.1 describes the data that will be needed for designing and/or analysing the results generated using the clinic-based method. For each variable, its use in the design (D) or analysis (A) phase is specified.

Table A2.1. Required clinic- and country-level data needed to implement the clinic-based acquired HIV drug resistance survey method outlined in section 2

Notation	Explanation	Required for design (D) or analysis (A) phase
N	Number of people (number of adults, if sampling for adult surveys, or number of children and adolescents, if sampling for children and adolescent surveys), receiving ART at the national level as used in the systematic sampling table	D,A
$prop_{DTG}$	National proportion of people receiving ART who are on DTG-containing regimens	D
C	Total number of clinics in the sampling table	D,A
c	Number of clinics to be sampled	D, A
N_k	Total number of people receiving ART from each sampled clinic, k *Also referred to as the clinic size as in the table used for systematic sampling*	D, A
$N_i{}^*{}_k$	Total number of people receiving regimen i (i=DTG or non-DTG) and *observed during the survey period*, from each sampled clinic, k	A

N_k: This number can be from the previous calendar year or for a recent 3-month period. The length of time is not important, provided it is consistent across all clinics and with the national number of people receiving ART (N). N_k is used to perform sampling of clinics, but the information is also necessary for the weights used in analysis.

A2.1 Determining the target sample sizes, overall and for individuals receiving DTG-containing regimens

The sample size calculations are conducted using the assumed values of the parameters specified in Table 1. Specifically, let p_{DTG}^{VS} denote the anticipated prevalence of viral suppression among individuals receiving DTG-containing regimens, assumed to be 0.9; let $p_{overall}^{VS}$ denote the overall anticipated prevalence of viral suppression among people on any ART regimen, assumed to be 0.85; and let L_{DTG} and $L_{overall}$ be the respective desired precision for the DTG and overall estimates, both set to 0.05. Unless otherwise specified, "individuals" and "people" refer to adults if the survey in adults is being conducted, or to children and adolescents if the survey in children and adolescents is being conducted.

To determine the target sample sizes, the following steps are taken:

1) Decide how many clinics will be sampled.
2) Calculate the effective sample sizes.
3) Adjust for clustering through an estimated intracluster correlation coefficient.
4) Incorporate inflations for viral load testing failure and the design effect due to imperfect weights.
5) Determine the per-clinic sample sizes for people receiving DTG-containing and non-DTG-containing regimens, incorporating the inflations, the finite population corrections, and the proportion of individuals on ART who are on DTG-containing regimens.

Annex 2, sections A2.1.1–A2.1.5 below describe the steps in detail. The first step of determining the number of clinics to sample requires an iterative process involving subsequent steps and is described in detail in Annex 2, section A2.1.5. For the purposes of illustrating the sample size calculation procedure, firstly begin by assuming the number of clinics to be sampled has already been determined and proceed directly with subsequent steps. Secondly, circle back to adjust and finalize the number of clinics to be sampled.

A2.1.1 Calculating effective sample sizes

The method for calculating a confidence interval for clustered surveys uses a t-distribution with df degrees of freedom, where the degrees of freedom are a function of the number of clinics to be sampled. Therefore, the effective sample size is also a function of the number of clinics to be sampled. When the design degrees of freedom are large (40 or greater), it is standard to use the normal approximation (z) instead of the t: $z_{0.975} \approx t_{df,0.975}$. However, since this design requires sampling a relatively small number of clinics, the design degrees of freedom will be small, and it is, thus, inadvisable to make this simplification. The consequence of using this simplification would be an underestimation of the total sample size required to achieve a given confidence interval half-width.

For a fixed number of clinics to be sampled, c, the effective sample sizes for estimating the prevalence of viral load suppression overall and the prevalence of viral load suppression for people receiving DTG-containing regimen is calculated first. The effective sample sizes (overall and DTG) refer to the number of people that would need to be sampled to achieve the desired precision for the overall and DTG estimates if a simple random sample were conducted. Each effective sample size is determined by the prevalence of the outcome and the desired width of the confidence interval and is then multiplied by three inflations, described in sections A2.1.2–A2.1.3, to yield the actual sample sizes for the survey.

The effective sample size formula is obtained by inverting the 95% Wald confidence interval with t-distribution defined by the design degrees of freedom. The design degrees of freedom are defined as df = (number of clinics to be sampled) - (number of strata) $= c - 2$, where the number of strata follows from people receiving either DTG-containing or non-DTG-containing regimens. Let $t_{c-2,0.975}$ denote the 0.975 quantile of a t-distribution with $c - 2$ degrees of freedom.

Prior to making any finite population corrections, the effective sample size for estimating the prevalence of viral suppression is given by the following formula:

$$n_{eff,i} = \left\lceil \frac{t_{c-2,0.975}^2 * p_i^{VS} * (1 - p_i^{VS})}{L_i^2} \right\rceil,$$

where $i = DTG$ indicates the estimate for people receiving DTG-containing regimens and $i = overall$ indicates the overall estimate for people receiving any regimen; $n_{eff,i}$ is the effective sample size for group i; p_i^{VS} is the expected prevalence of viral suppression for group i; L_i is the desired absolute precision for group i and $\lceil \cdot \rceil$ is the ceiling function and rounds the inner value up to the nearest integer. The effective sample sizes are a function of the number of clinics to be sampled, c.

A2.1.2 Adjusting sample sizes for clustering

Because clinics ("clusters") are sampled first and then individuals within clinics are sampled, the effective sample sizes (overall and DTG) must be adjusted to account for increased variance due to expected clustering of the outcome within clinics. People from the same clinic may have more similar viral load outcomes than people from different clinics in the same country. This adjustment is made by setting an anticipated level for the ICC, which measures the similarity of viral suppression among people within the same clinic. A higher ICC value accounts for a higher level of clustering.

A2.1.2.1 Options for adjustment of clustering

Two design options are available to countries when choosing how to adjust for clustering in the viral suppression estimates. The options differ in the value of the anticipated ICC, which corresponds to a difference in the level of clustering that is accounted for in the design.

Option 1 sets the ICC to be 0.09, which is the median of the estimated country-level ICC values from ADR surveys implemented during the period 2016–2019 (see Annex 2, section A2.1.2.2 for more details). This option accounts for higher levels of clustering but, as a result, requires more clinics to be included as well as larger overall sample sizes.

In contrast, Option 2 sets the ICC to be 0.06, which is the 25th percentile of the estimated country-level ICC values from ADR surveys implemented during the period 2016–2019. This option accounts for a lower level of clustering in the design and, as a result, requires fewer clinics and smaller sample sizes. In the WHO online sample size calculator (available at https://worldhealthorg.shinyapps.io/ADR_ClinicBasedMethod/) countries can choose whether or not Option 1 or 2 will be used depending on whether the minimum number of clinics displayed can be sampled.

Note that, in reality, the true ICC for the viral suppression prevalence in a country is unknown; the anticipated ICC is our best estimate based on historical data. Setting the ICC to the 25th percentile provides design options that are more feasible for some countries. However, it is possible that this ICC is too low compared with the true ICC. In this case, the resulting estimates and confidence intervals will still be correctly estimated; however, the confidence intervals will likely be wider than the targeted ±5% precision.

A2.1.2.2 Estimating the intracluster correlation coefficient

Estimation of the ICC for prevalence of viral load suppression was conducted by using ADR survey data that was shared with WHO to inform the global HIV drug resistance reports. A total of 20 datasets from ADR surveys among adults that were carried out by 10 countries were used. For each survey carried out in each country, the ICC was calculated using the following procedure:

1) Virally suppressed was defined as a viral load below 1000 copies/mL.
2) A binary variable (labelled "VL_SUPPRESSED_BN") was generated indicating whether a patient was virally suppressed (1 if suppressed; 0 if not suppressed; missing if no specimen was collected or specimen was collected but no data were available).
3) A variable (labelled "SITE_ID") was generated listing the unique site IDs of the clinics that participated in each survey.
4) In Stata 15.1 the ICC was calculated using the following code:
   melogit VL_SUPPRESSED_BN || SITE_ID:
   estat icc

The ICC analysis is summarized in Table A2.2. The median ICC calculated was 0.093. The first quartile (25th percentile) ICC value was 0.065 and the third quartile (75th percentile) ICC value was 0.157. The ICC analysis was not feasible with 5 datasets. ADR12 and ADR48 correspond to the surveys among adults receiving ART for at least 12 and 48 months, respectively, detailed in the 2014 WHO ADR protocol. Since time points are not specified for this protocol, ICC data from both time points are used and considered equally to estimate the median ICC. The analysis of the ICC for the outcome of the prevalence of acquired HIV drug resistance was not feasible given the small number of observations (genotypes available) per clinic.

Table A2.2. Analysis of the intracluster correlation coefficient of the acquired HIV drug resistance surveys among adults informing the global HIV drug resistance reports

Country	Survey	Number of clinics	ICC	95% confidence interval		Comments
				LCL	UCL	
El Salvador	ADR12	12	0.0212	0.0000331	0.9341533	
El Salvador	ADR48	12	0.0805842	0.0181693	0.2933473	
Eswatini	ADR12	26	0.0812238	0.010466	0.424931	
Eswatini	ADR48	21	0.0760926	0.0066461	0.5034328	
Guatemala	ADR12	10	0.1521949	0.0371885	0.4548437	
Guatemala	ADR48	10	0.3433349	0.1489039	0.6097565	
Honduras	ADR12	20				Analysis not feasible
Honduras	ADR48	22				Analysis not feasible
Myanmar	ADR12	26				Analysis not feasible
Myanmar	ADR48	19	0.0929436	0.0171538	0.3756181	
Nicaragua	ADR12	19				Analysis not feasible
Nicaragua	ADR48	17	0.0531482	0.0018734	0.6266874	
Senegal	ADR12	29	0.1148715	0.0236182	0.4104752	
Senegal	ADR48	30	0.1608316	0.0329606	0.5186959	
South Sudan	ADR12	14	0.110851	0.0308039	0.3284219	
South Sudan	ADR48	9	0.0189659	0.0018084	0.1710149	
Uganda	ADR12	23				Analysis not feasible
Uganda	ADR48	23	0.0331832	0.0063834	0.1549522	
Viet Nam	ADR12	30	0.2042811	0.0379225	0.6257563	
Viet Nam	ADR48	25	0.2950798	0.1044516	0.6003766	
First quartile (25th percentile)			0.0646204			
Second quartile (median ICC)			0.0929436			
Third quartile (75th percentile)			0.15651325			

ICC=intracluster correlation coefficient; LCL=lower confidence limit; UCL=upper confidence limit
Note: ADR12 refers to WHO's 2014 acquired drug resistance survey method assessing viral load suppression in people on ART for 12±3 months. ADR 48 refers to WHO's 2014 acquired drug resistance survey method assessing viral load suppression in people on ART for 48 or more months.

Stata code to support the calculation of the actual ICC during the analysis phase is provided in Annex 2, section A2.4, and future protocols may update the ICC based on these estimates.

A2.1.3 Sample size inflation for viral load testing failure and imperfect weights

After adjusting the effective sample sizes (overall and DTG) for the clustering effect, the sample sizes must undergo two additional inflations. The first additional inflation accounts for the viral load testing failure rate among specimens sent to laboratories to receive viral load testing. Since not all specimens will have a viral load test result to contribute to analysis, WHO recommends incorporating an anticipated viral load testing failure rate, f, of 10% into sample size calculations.

The second additional inflation adjusts for sampling proportional to proxy size, rather than directly proportional to size. Ideally, countries would sample clinics proportional to the number of individuals on DTG-containing regimens and the number on non-DTG-containing regimens. However, this level of granular detail is generally unavailable in most countries. Thus, it is recommended that countries use PPPS sampling, where the national proportion of individuals on DTG-containing regimens is combined with the clinic-specific number of individuals on ART based on historical data to approximate the clinic-specific number of individuals on DTG- and non-DTG-containing regimens.

Because a proxy measure is used for sampling, updated weights will need to be included in the analysis. If the actual number of individuals currently on ART differs from historical numbers or if the proportion of people receiving DTG-containing regimens varies widely by clinic, then the design and weights could be far from proportional, thereby resulting in increased variances and a loss of efficiency. The sample size is inflated to offset this increased variance by incorporating the design effect for imperfect weighting (DE_{info}). For this activity, $DE_{info} = 1.5$ as used in the prior surveys, which corresponds to inflating the sample size by 50% to account for the imperfect information. As surveys are implemented, it is recommended that the design effect for imperfect weighting be re-evaluated and adjusted as necessary for future iterations of the survey.

A2.1.4 Determining country-specific per-clinic sample sizes by regimen

The country-specific per-clinic sample sizes for people receiving DTG- and non-DTG-containing regimens can be calculated by incorporating the finite population corrections, the proportion of individuals on ART who are on DTG-containing regimens and the inflations described in Annex 2, sections A2.1.2–A2.1.3.

Country-specific sample sizes will vary depending on 1) the number of clinics in the country, C; 2) the number of people on ART in the country, N; 3) the proportion of individuals on DTG-containing regimens, $prop_{DTG}$; and 4) the amount of clustering that a country can afford to accommodate in the survey design, represented through the ICC (as described in Annex 2, section A2.1.2). The country-specific per-clinic sample size for group i (i = DTG or overall) is given by:

$$m_i = \frac{1 - ICC}{\left(\left(\frac{n}{DE_{info} * n_{eff,i}}\right) - ICC\left(1 - \frac{c}{C}\right) + \frac{C}{N_i}(1 - ICC)\right)0.9}.$$

Note that $N_{overall} = N$ and $N_{DTG} = N * prop_{DTG}$.

Depending on the proportion of individuals receiving ART who are taking DTG-containing regimens, the overall sample size may be larger than the DTG sample size to the extent that $m_{overall} * prop_{DTG} > m_{DTG}$. This is particularly the case when the proportion on DTG-containing regimens is small, such as 10%. In order to ensure enough specimens from people receiving DTG-containing regimens are sampled to meet the sample size requirements for both the DTG and overall estimates, the per-clinic sample size for people receiving DTG-containing regimens is given by:

$$\tilde{m}_{DTG} = \max\{m_{DTG}, m_{overall} * prop_{DTG}\}$$

The per-clinic sample size for people receiving non-DTG-containing regimens is given by

$$m_{nonDTG} = m_{overall} * (1 - prop_{DTG}).$$

The total per-clinic sample size will be the sum of the per-clinic DTG and per-clinic non-DTG sample sizes: $\tilde{m}_{DTG} + m_{nonDTG}$. Multiplying the per-clinic sample size by the number of clinics to be sampled gives the total sample size:

$$M_{total} = (\tilde{m}_{DTG} + m_{nonDTG}) * c.$$

The above sample size calculations are automated in the WHO online sample size calculator (available at https://worldhealthorg.shinyapps.io/ADR_ClinicBasedMethod/), in which countries can enter their relevant data and clustering preference and obtain all target sample sizes as outputs.

A2.1.5 Determining the number of clinics to sample

Sampling more clinics is generally preferred from a statistical perspective, although this often results in increased costs. When choosing the number of clinics to be sampled, c, one key constraint is the minimum number of clinics that are necessary to ensure that the per clinic sample size per regimen is greater than zero. That is, for i = DTG or overall,

$$m_i = \frac{1 - ICC}{\left(\left(\frac{c}{DE_{info} n_{eff,i}}\right) - ICC\left(1 - \frac{c}{C}\right) + \frac{C}{N_i}(1 - ICC)\right) 0.9} > 0.$$

Here, m_i is the per clinic sample size for group i (i = DTG or overall), ICC is the estimated value for the intracluster correlation coefficient, DE_{info} is the design effect due to imperfect weighting, $n_{eff,i}$ is the effective sample size for group i (see Annex 2, section A2.1.1), c is the number of clinics to be sampled, C is the total number of clinics in the country, N_i is the total number of individuals in group i in the country, and 0.9 accounts for the viral load testing failure rate of 10%.

Rearranging the above equation and using the fact that DE_{info}, $n_{eff,i}$, ICC, and M_i are all greater than zero, the following inequality is obtained:

$$c > \frac{(C + N_i)ICC - C}{N_i\left(\frac{1}{DE_{info} n_{eff,i}} + \frac{ICC}{C}\right)},$$

which provides a lower bound on the allowable values for the number of clinics to be sampled, c. This condition specifies the minimum number of clinics necessary to ensure the per clinic sample sizes are valid. However, notice that the right-hand side of the inequality is also an implicit function of c through $n_{eff,i}$ because the t-distribution degrees of freedom in the effective sample size depend on c. To solve the above equation for the value of c, one can start at two and iterate through increasing values of c until the inequality above is satisfied.

The WHO online sample size calculator (available at https://worldhealthorg.shinyapps.io/ADR_ClinicBasedMethod/) implements this approach with an additional condition requiring the total sample size to be less than or equal to 1500, the threshold of feasibility determined based on previous ADR survey implementations. This additional condition gives rise to a minimum number of clinics that is slightly larger than what is required by the inequality above but is preferable because it ensures that the total sample size will be feasible. Note that as the number of clinics to be sampled increases, the total target sample size for the survey will decrease. This occurs because: 1) sampling more clinics reduces the design effect by allowing for a more representative sample of clinics, and 2) sampling more clinics increases the degrees of freedom in the t-distribution, which lowers the value of the 0.975 quantile and consequently decreases the effective sample size.

A2.2 Anticipated precision in the ADR estimates

The sample sizes for this activity are set to achieve precision in viral suppression estimates. However, a key secondary outcome is precision in the resulting ADR estimates, specifically, the prevalence of ADR overall and the prevalence of DTG-specific ADR in individuals on DTG-containing regimens. The WHO sample size calculator (available at https://worldhealthorg.shinyapps.io/ADR_ClinicBasedMethod/) reports the anticipated precision in these estimates, based on the following assumptions:

- 10% of samples will not successfully test for viral load: $f = 0.1$.
- Of those tested, the viral suppression rate is expected to be 85% overall and 90% in DTG-specific populations: $p^{VS}_{overall} = 0.85$ and $p^{VS}_{DTG} = 0.9$.
- Of those with viral non-suppression (VNS), the genotypic testing failure rate is 30%: $g = 0.3$.
- Given the current absence of information on an appropriate ICC for ADR estimates, the variance of the estimate is inflated by a flat design effect of 1.5 to account for cluster sampling. This is denoted by DE=1.5.
- The expected ADR prevalence among people with VNS is 50% overall and the DTG-specific ADR prevalence among those on DTG-containing regimens is 3.5%: $p^{ADR}_{overall} = 0.5$ and $p^{ADR}_{DTG} = 0.035$.

The anticipated available sample size for the ADR estimates for group i (i=DTG or overall) will consist of the per clinic sample sizes multiplied by the number of clinics sampled, one minus the viral load testing failure rate, one minus the expected viral suppression rate, and one minus the genotyping testing failure rate:

$$M_i^{ADR} = m_i * c * (1 - f) * (1 - p_i^{VS}) * (1 - g).$$

This restricts the estimated ADR sample size to the anticipated number of specimens that are sampled, undergo successful viral load testing, are virally non-suppressed and undergo successful genotyping.

Because of the clustering design and the DE=1.5 adjustment, the effective sample size is

$$n_{eff,i}^{ADR} = \frac{M_i^{ADR}}{DE} = \frac{M_i^{ADR}}{1.5}.$$

Therefore, the precision of the ADR estimate in group i will be

$$\pm t_{c-2,0.975} * \sqrt{\frac{p_i^{ADR}(1 - p_i^{ADR})}{n_{eff,i}^{ADR}}}.$$

A2.3 Analysis

This section describes the statistical details for each of the five outcomes. Stata code is provided to automate this analysis. Unless otherwise specified, "individuals" refers to adults if the survey in adults is being conducted, or to children and adolescents if the survey in children and adolescents is being conducted.

A2.3.1 Notation

The following notation is used throughout this section.
i=subscript for regimen, where i=DTG or non-DTG
k=subscript for clinic
l=subscript for individuals
a=subscript for membership in the subgroup of interest
I=set consisting of the two regimens: {DTG, non-DTG}
C=total number of clinics
c=total number of clinics to be sampled (not the number of unique clinics sampled)
N=estimated total number of people receiving ART used in the sampling frame
N_k=estimated number of people receiving ART at clinic k used in the sampling frame
$N_{i,k}^*$=number of individuals receiving regimen i in clinic k, observed during the survey period
$m_{i,k}^*$=number of sampled individuals receiving regimen i in clinic k with successful viral load tests
$m_{i,k}^{VNS}$=number of sampled individuals receiving regimen i in clinic k with VNS
$m_{i,k}^{VNS,geno}$=number of sampled individuals receiving regimen i in clinic k with VNS and successful genotyping
N_i^*=number individuals receiving regimen i observed during the survey period, summed over all sampled clinics
N_i^{VNS}=best estimate of the total number of individuals on regimen i with VNS
m_i^{VNS}=number of sampled individuals with VNS on regimen i observed during the survey period, summed over all sampled clinics
c_{adults}=number of clinics to be sampled for the adult survey
N_{adults}=total number of adults on ART
$N_{k,adults}$=number of adults on ART at clinic k
c'_{child}=number of clinics to be sampled for the children and adolescent survey minus the number of clinics already included via the adult sampling process
N'_{child}=total number of children and adolescents on ART minus the total number in clinics included via the adult sampling process
$N_{k,child}$=number of children and adolescents receiving ART at clinic k
$Y_{i,k,l}$=binary outcome of viral suppression for specimen l on regimen i in clinic k
$X_{i,k,l}$=binary outcome variable for acquired HIV drug resistance for specimen l from individuals with VNS on regimen i in clinic k
$\hat{p}_{overall}^{VS}$=estimated prevalence of viral suppression among all individuals
$\hat{p}_{DTG}^{VS}$=estimated prevalence of viral suppression among people receiving DTG-containing regimens
$\hat{p}_{i,k}^{VS}$=estimated prevalence of viral suppression among individuals receiving regimen i in clinic k
$\hat{p}_i^{VS}$=estimated prevalence of viral suppression among individuals receiving regimen i
$\hat{p}_a^{VS}$=estimated prevalence of viral suppression among individuals in subgroup of interest
$\hat{p}_{overall}^{ADR}$=estimated prevalence of any ADR among specimens with VNS
$\hat{p}_{DTG}^{ADR}$=estimated prevalence of DTG-specific ADR among specimens with VNS from people receiving DTG-containing regimens

A2.3.2 Calculation of survey sampling weights

A2.3.2.1 Calculation of survey sampling weights for the survey in adults

Survey sampling weights can be calculated by taking the reciprocal of the probability of an individual being selected. For the survey design, sampling unfolds in two stages.

- The first stage involves clinics as the primary sampling units. c clinics are sampled using PPPS sampling, where sampling occurs proportional to the number of adults on ART at the clinic, N_k. This number, c, is the pre-specified number of clinics serving adults to be sampled, not the number of unique clinics selected, as some clinics may be sampled more than once.
- The second stage involves stratification by regimen, with specimens as the secondary sampling units. Regimen i refers to either DTG-containing regimens or non-DTG-containing regimens. For each regimen i and clinic k, $m^*_{i,k}$ adult specimens with successful VL tests and complete desired variables ultimately contribute to the observed sample size out of a total of $N^*_{i,k}$ adults observed during the 3-month survey period.

The probability of an adult specimen being sampled is given by the product of the clinic sampling probability and the patient sampling probability:

$$\begin{aligned} &P(\text{specimen } l \text{ on regimen } i, \text{ from clinic } k, \text{ is selected}) \\ &= P(\text{clinic } k \text{ is selected}) * P(\text{specimen } l \text{ on regimen } i \text{ is selected} \mid \text{clinic } k \text{ is selected}) \\ &= \frac{cN_k}{N} \times \frac{m^*_{i,k}}{N^*_{i,k}} \end{aligned}$$

The survey sampling weight for a specimen l on regimen i from clinic k is the inverse of the probability of selection:

$$w_{i,k,l} = \frac{N}{cN_k}\frac{N^*_{i,k}}{m^*_{i,k}}.$$

A2.3.2.2 Calculation of survey sampling weights for the survey in children and adolescents

Survey type: children and adolescents separate implementation

If the survey among children and adolescents is implemented separately from the survey among adults, all sample size and analysis calculations will be equivalent to those described for the survey among adults. The only change will be that the individuals included in the survey are children and adolescents rather than adults.

Survey type: children and adolescents combined implementation

If the survey among children and adolescents is implemented alongside the survey among adults, as detailed in section 2.6, survey sampling weights are still calculated by taking the reciprocal of the probability of an individual being selected; however, this probability is complicated due to clinics serving both adults and children and adolescents and selected for the adult survey being retained for inclusion in the child/adolescent survey as well. As a result, for each clinic sampled for the survey among children and adolescents, the clinic sampling probability differs depending on whether the clinic serves adults and children and adolescents or whether it serves children and adolescents only.

For a clinic k serving both adults and children and adolescents, the clinic sampling probability is as follows. Let S_c denote the sample for children and adolescents and let S_a denote the sample for adults. The probability of clinic k being selected for the child/adolescent sample is,

$$\begin{aligned} P(\text{clinic } k \text{ is in } S_c) &= P(\text{clinic } k \text{ is in } S_c \mid \text{clinic } k \text{ is in } S_a) \times P(\text{clinic } k \text{ is in } S_a) \\ &\quad + P(\text{clinic } k \text{ is in } S_c \mid \text{clinic } k \text{ is not in } S_a) \times P(\text{clinic } k \text{ is not in } S_a) \\ &\approx 1 \times \left(\frac{N_{k,adults}c_{adults}}{N_{adults}}\right) + \left(\frac{N_{k,child}c'_{child}}{N'_{child}}\right) \times \left(1 - \frac{N_{k,adults}c_{adults}}{N_{adults}}\right), \end{aligned}$$

where, for the adult sampling process, c_{adults} is the number of clinics to be sampled for the adult survey, N_{adults} is the total number of adults receiving ART, and $N_{k,adults}$ is the number of adults receiving ART at clinic k. Similarly, for the subsequent child/adolescent sampling process, c'_{child} is the number of clinics to be sampled for the child/adolescent survey minus the number of clinics already included via the adult sampling process, N'_{child} is the total number of children and adolescents on ART minus the total number in clinics included via the adult sampling process, and $N_{k,child}$ is the number of children and adolescents receiving ART at clinic k.

For a clinic k serving only children and adolescents, the clinic sampling probability for the child/adolescent sample is

$$\begin{aligned} P(\text{clinic } k \text{ is in } S_c) &= P(\text{clinic } k \text{ is in } S_c \mid \text{clinic } k \text{ is in } S_a) \times P(\text{clinic } k \text{ is in } S_a) \\ &\quad + P(\text{clinic } k \text{ is in } S_c \mid \text{clinic } k \text{ is not in } S_a) \times P(\text{clinic } k \text{ is not in } S_a) \\ &\approx 0 + \left(\frac{N_{k,child} c'_{child}}{N'_{child}}\right) \times 1. \end{aligned}$$

Note that these are closed-form approximations that are useful in practice, but complete theoretical results remain to be developed.

Next, the above clinic sampling probabilities are combined with the individual within-clinic sampling probability to obtain the full survey sampling weights. Recall that sampling unfolds in two stages.

- The first stage involves clinics as the primary sampling units. As was described in section 2.6, clinics sampled for the adult survey and serving both adults and children and adolescents are automatically included in the sample. Then c'_{child} additional clinics are sampled using PPPS sampling, where sampling occurs proportional to the number of children and adolescents on ART at the clinic, $N_{k,child}$.
- The second stage involves stratification by regimen, with individual specimens as the secondary sampling units. Regimen i refers to either DTG-containing regimens or non-DTG-containing regimens. For each regimen i and clinic k, $m^*_{i,k}$ specimens with successful viral load tests and complete desired variables ultimately contribute to the observed sample size out of a total of $N^*_{i,k}$ individuals observed during the 3-month survey period.

The probability of a specimen being sampled is given by the product of the clinic sampling probability and the patient sampling probability:

$$\begin{aligned} &P(\text{specimen } l \text{ on regimen } i, \text{ from clinic } k, \text{ is selected}) \\ &= P(\text{clinic } k \text{ is selected}) \times P(\text{specimen } l \text{ on regimen } i \text{ is selected} \mid \text{clinic } k \text{ is selected}) \\ &\approx \begin{cases} \left[\left(\frac{N_{k,adults} c_{adults}}{N_{adults}}\right) + \left(\frac{N_{k,child} c'_{child}}{N'_{child}}\right)\left(1 - \frac{N_{k,adults} c_{adults}}{N_{adults}}\right)\right] \times \frac{m^*_{i,k}}{N^*_{i,k}}, & \text{if clinic } k \text{ serves adults and children} \\ \left(\frac{N_{k,child} c'_{child}}{N'_{child}}\right) \times \frac{m^*_{i,k}}{N^*_{i,k}}, & \text{if clinic } k \text{ serves only children.} \end{cases} \end{aligned}$$

The survey sampling weight for a specimen l on regimen i from clinic k is the inverse of the probability of selection:

$$w_{i,k,l} = \frac{1}{P(\text{specimen } l \text{ on regimen } i, \text{ from clinic } k, \text{ is selected})}.$$

A2.3.3 Analysis for outcomes 1 and 2

Outcome 1 is defined as: prevalence of viral suppression among individuals receiving ART, regardless of the ART regimen.
Outcome 2 is defined as: prevalence of viral suppression among individuals receiving a DTG-containing regimen.

Outcome1: Prevalence and variance of viral suppression among individuals receiving ART, regardless of the ART regimen

Prevalence

An unbiased estimate of the prevalence of viral suppression amongst all individuals on ART can be obtained by using a ratio estimator. The numerator is an estimate of the total number of people in the country achieving viral load suppression. The denominator is an estimate of the total number of people receiving ART in the country.

As defined in Annex 2, section A2.3.2, the sampling weight for each specimen among individuals receiving regimen i (i =DTG or non-DTG) is given by $w_{i,k,l}$. Let $Y_{i,k,l}$ denote the binary outcome variable for viral suppression, equal to 1 if patient l on regimen i in clinic k is virally suppressed and 0 if not. Let $\hat{p}_{i,k}^{VS}$ be the estimated prevalence of viral suppression among individuals receiving regimen i in clinic k, and let I denote the set of possible regimens: {DTG, non-DTG}. The prevalence estimate for viral suppression is given by

$$\hat{p}_{overall}^{VS} = \frac{\sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^{*}} w_{i,k,l}Y_{i,k,l}}{\sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^{*}} w_{i,k,l}}.$$

Variance

The variance formula is obtained by linearizing the ratio estimator. Let $\hat{p}_{overall}^{VS} = \frac{\hat{z}}{\hat{v}}$, where $\hat{z} = \sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^{*}} w_{i,k,l}Y_{i,k,l}$ and $\hat{v} = \sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^{*}} w_{i,k,l}$. The linearized variance estimator is given by

$$\widehat{Var}(\hat{p}_{overall}^{VS}) = \frac{1}{\hat{v}^2}\left(\hat{p}_{overall}^{VS\,2}\widehat{Var}(\hat{v}) + \widehat{Var}(\hat{z}) - 2\hat{p}_{overall}^{VS}\widehat{Cov}(\hat{z},\hat{v})\right),$$

where $\widehat{Var}(\hat{v}), \widehat{Var}(\hat{z})$, and $\widehat{Cov}(\hat{z},\hat{v})$ can be estimated by the corresponding sample variances and covariances, which are also obtained through linearization and incorporate the finite population corrections at both stages of sampling. A 95% confidence interval can be calculated using a standard Wald formula or a Wald formula transformed to the logit scale (default in Stata).

Outcome 2: Prevalence and variance of viral suppression among individuals receiving DTG-containing regimens

Prevalence

An unbiased estimate of viral suppression prevalence among individuals receiving DTG-containing regimens is the same as the overall estimate, but it only uses data from individuals receiving DTG-containing regimens. The sampling weight for each specimen is given by $w_{DTG,k,l}$. Let $Y_{DTG,k,l}$ denote the binary outcome variable for viral suppression among individuals receiving DTG-containing regimens, equal to 1 if patient l in clinic k is virally suppressed and 0 if not. Similarly, let $\hat{p}_{DTG,k}^{VS}$ be the estimated prevalence of viral suppression among individuals receiving DTG-containing regimens in clinic k. The prevalence estimate is given by

$$\hat{p}_{DTG}^{VS} = \frac{\sum_{k=1}^{c}\sum_{l=1}^{m_{DTG,k}^{*}} w_{DTG,k,l}Y_{DTG,k,l}}{\sum_{k=1}^{c}\sum_{l=1}^{m_{DTG,k}^{*}} w_{DTG,k,l}}.$$

Variance

The variance formula is obtained by linearizing the ratio estimator for stratified two-stage cluster samples. For simplicity, let $\hat{p}_{DTG}^{VS} = \frac{\hat{z}}{\hat{v}}$, where $\hat{z} = \sum_{k=1}^{c}\sum_{l=1}^{m_{DTG,k}^{*}} w_{DTG,k,l}Y_{DTG,k,l}$ and $\hat{v} = \sum_{k=1}^{c}\sum_{l=1}^{m_{DTG,k}^{*}} w_{DTG,k,l}$. The linearized variance estimator is given by

$$\widehat{Var}(\hat{p}_{DTG}^{VS}) = \frac{1}{\hat{v}^2}\left(\widehat{Var}(\hat{z}) + \hat{p}_{DTG}^{VS\,2}\widehat{Var}(\hat{v}) - 2\hat{p}_{DTG}^{VS}\widehat{Cov}(\hat{z},\hat{v})\right),$$

where $\widehat{Var}(\hat{v}), \widehat{Var}(\hat{z})$, and $\widehat{Cov}(\hat{z},\hat{v})$ can be estimated by the corresponding sample variances and covariances, which are also obtained through linearization and incorporate the finite population corrections at both stages. A 95% confidence interval can be calculated using a standard Wald formula or a Wald formula transformed to the logit scale (default in Stata).

A2.3.4 Analysis for outcome 3

Outcome 3 is defined as: prevalence of viral suppression among individuals receiving ART, stratified by age band, gender, ART regimen (DTG-containing versus non-DTG-containing regimen), ART line, previous ART regimen, pregnancy status and breastfeeding status, if known.

Subgroup analysis can be performed to estimate the prevalence of viral suppression in subgroups determined by other key variables such as age, gender, ART regimen, ART line, previous ART regimen, pregnancy status and breastfeeding status. Each of these analyses will involve a complete-case analysis that excludes missing data. Those in the subpopulation and not missing the variable of interest are used to calculate the prevalence estimate, and all individuals are used to calculate the standard errors. This allows the randomness of the subgroup size in the sample to be incorporated into variance estimation.

The methods presented in this section pertain to subgroups where membership is not defined by a stratifying variable. For subgroup analysis in which membership in the subgroup is defined by the stratifying variable of regimen (DTG or non-DTG), then analysis can proceed by restricting it to the subgroup of interest and using the outcome relevant to the type of resistance desired.

Prevalence

Prevalence estimates for the subgroup of interest involve a ratio estimator: the denominator is the number of sampled individuals belonging to the subgroup, and the numerator is the number of sampled individuals belonging to the subgroup with viral suppression.

The strata are defined by regimen: i =DTG or non-DTG. Suppose the subgroup of interest is denoted by a, with membership in the subgroup labelled through an indicator variable, $I_{i,k,l}(a)$, equal to 1 if individual l on regimen i in clinic k belongs in the subgroup, and equal to 0 otherwise. The estimated prevalence of viral suppression among those within the subgroup is:

$$\hat{p}_a^{VS} = \frac{\sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^*} I_{i,k,l}(a) w_{i,k,l} Y_{i,k,l}}{\sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^*} I_{i,k,l}(a) w_{i,k,l}},$$

where $Y_{i,k,l}$ is the binary outcome variable for viral suppression for individual l on regimen i in clinic k.

Variance

Variance estimation for the subgroup prevalence is approximated using a linearized variance estimator. The denominator in the prevalence estimate is not fixed because the number of sampled individuals belonging to the subgroup is random and depends on the specific sample that is chosen. To account for this additional sample-to-sample variability, linearized variance estimation incorporates all observations but multiplies the weights of those not in the subgroup by 0. Note that the validity of linearization depends on having a large enough size of the subgroup in the sample. If the subgroup sample size is too small, composite or model-based estimators may be considered, or it may be decided not to conduct the subgroup analysis.

For simplicity, let $\hat{p}_a^{VS} = \frac{\hat{z}}{\hat{v}}$, where $\hat{z} = \sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^*} I_{i,k,l}(a) w_{i,k,l} Y_{i,k,l}$ and $\hat{v} = \sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^*} I_{i,k,l}(a) w_{i,k,l}$. The linearized variance estimator is given by

$$\widehat{Var}(\hat{p}_a^{VS}) = \frac{1}{\hat{v}^2}\left(\hat{p}_a^{VS^2}\widehat{Var}(\hat{v}) + \widehat{Var}(\hat{z}) - 2\hat{p}_a^{VS}\widehat{Cov}(\hat{z},\hat{v})\right),$$

where $\widehat{Var}(\hat{v}), \widehat{Var}(\hat{z})$, and $\widehat{Cov}(\hat{z},\hat{v})$ can be estimated by the corresponding sample variances and covariances, which are also obtained through linearization and incorporate the finite population correction at both stages of sampling. A 95% confidence interval can be calculated using a standard Wald formula or a Wald formula transformed to the logit scale (default in Stata).

These methods are given for overall prevalence and variance estimates of resistance among a subgroup of interest. For subgroup estimates restricted to a certain regimen, simply restrict the data first, then apply these estimation and inference methods.

A2.3.5 Analysis for outcomes 4 and 5

Outcome 4 is defined as: prevalence of acquired HIV drug resistance among individuals with VNS and receiving ART, regardless of the ART regimen.

Outcome 5 is defined as: prevalence of acquired HIV drug resistance to DTG among individuals with VNS and receiving ART who are taking a DTG-containing regimen.

Outcome 4: Prevalence and variance of acquired HIV drug resistance among individuals with VNS and receiving ART, regardless of the ART regimen

Prevalence

Prevalence of acquired HIV drug resistance will be estimated using sampled specimens that have undergone successful viral load testing, have VNS, and have undergone successful genotyping testing. Due to a low anticipated rate of VNS and considerable failure rates for viral load and genotyping testing, the number of sampled specimens per clinic satisfying these requirements is likely to be only one or two specimens, if any. As a result, true inverse probability sampling weights will likely lead to unstable estimates with high variability.

To counteract this, estimation will proceed via a partially weighted analysis, where clinic sampling weights will be ignored, and data will only be weighted to account for stratification by regimen and nonresponse from genotyping failure. This may result in increased bias, but the benefits of decreasing variability in the resulting estimates are substantial given the particularly small sample sizes. Countries have not been advised to inflate sample sizes for estimated prevalence of acquired HIV drug resistance because the outcomes are not of primary interest, and the sample size increases necessary would result in surveys with infeasible sample sizes.

For the prevalence estimate, the numerator is an estimate of the total number of individuals with VNS and with ADR, and the denominator is an estimate of the total number of individuals with VNS. Stratification by regimen is accounted for but not at the clinic level; instead, we use a stratification weight defined as the estimated total number of people receiving regimen i with VNS divided by the number of sampled individuals receiving regimen i with VNS. Let $N_i^* = \sum_{k=1}^{c} N_{i,k}^*$ be the number individuals on regimen i observed during the survey period, summed over all sampled clinics. The best estimate of the total number of individuals on regimen i with VNS is given by $N_i^{VNS} = N_i^*(1 - \hat{p}_i^{VS})$, where $\hat{p}_i^{VS}$ is the regimen-specific viral suppression estimate ($\hat{p}_{DTG}^{VS}$ is calculated for Outcome 2 and $\hat{p}_{nonDTG}^{VS}$ can be obtained similarly) and is treated as a constant. In addition, let $m_i^{VNS} = \sum_{k=1}^{c} m_{i,k}^{VNS}$ be the total number of sampled individuals with VNS on regimen i observed during the survey period, summed over all sampled clinics. Then, for each regimen i, the weight that accounts for stratification by regimen is given by $\frac{N_i^{VNS}}{m_i^{VNS}}$. This weight avoids the instability of clinic-specific weights while still making some adjustment to reflect the differential proportions of the two regimen types in the estimates and the anticipation that ADR estimates may differ between DTG and non-DTG regimens.

The non-response sampling weight is defined, for each clinic k and regimen i, as the number of sampled individuals with VNS divided by the number of sampled individuals with VNS and successful genotyping: $\frac{m_{i,k}^{VNS}}{m_{i,k}^{VNS,geno}}$. The non-response sampling weight is missing for individuals with VNS and missing genotype and is equal to 1 for individuals with viral suppression. The non-response weight assumes that genotyping failure is unrelated to the presence of ADR. Since the population is restricted to individuals with VNS, this is a subpopulation analysis and individuals with viral suppression or missing genotype will not contribute to the estimate but will contribute to the variance.

The final weight for patient l on regimen i in clinic k is given by $v_{i,k,l} = \frac{N_i^{VNS}}{m_i^{VNS}} \cdot \frac{m_{i,k}^{VNS}}{m_{i,k}^{VNS,geno}}$. In addition, let $X_{i,k,l}$ denote the binary outcome variable for acquired HIV drug resistance among people with VNS, equal to 1 if patient l on regimen i in clinic k is resistant and 0 if not. The estimate for prevalence of acquired HIV drug resistance among individuals with VNS, regardless of ART regimen, is given by

$$\hat{p}_{overall}^{ADR} = \frac{\sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^{VNS,geno}} v_{i,k,l}X_{i,k,l}}{\sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m_{i,k}^{VNS,geno}} v_{i,k,l}} = \frac{\sum_{k=1}^{c}\sum_{i\in I}\frac{N_i^{VNS}}{m_i^{VNS}}\cdot\frac{m_{i,k}^{VNS}}{m_{i,k}^{VNS,geno}}\sum_{l=1}^{m_{i,k}^{VNS,geno}} X_{i,k,l}}{\sum_{k=1}^{c}\sum_{i\in I}\frac{N_i^{VNS}}{m_i^{VNS}}\cdot m_{i,k}^{VNS}}.$$

Variance

Given the small sample sizes available for estimation of acquired HIV drug resistance, a cluster bootstrap method rather than linearization is recommended for variance estimation. The bootstrap method involves generating a large number of independent resamples (that is, bootstrap samples), each drawn with replacement from the original sample, to mimic the original sampling procedure and obtain approximate sampling distributions of the prevalence estimates. To address the clustering effect of ADR within clinics, the method resamples clinics rather than individual specimens. $c - 1$ clinics are resampled to ensure efficiency of bootstrap estimators without violating the natural parameter ranges.[1] For each of the bootstrap samples, the prevalence estimate specified above is calculated.

To obtain a bootstrap variance estimator that is adjusted for bias, an internal rescaling procedure is applied to the stratification weights to produce bootstrap replicate weights.[2,3] The adjusted weight for replicate r is given by

$$v_{i,k,l}^{(r)} = \frac{c}{c-1} r_k v_{i,k,l}, r = 1, \ldots, R$$

where r_k is the number of times cluster k was resampled and R is the total number of bootstrap samples. These replicate weights represent the resamples and also allow variance estimates to be calculated with only knowledge of the first stage of sampling.

[1] Kolenikov S. Resampling variance estimation for complex survey data. Stata J. 2010;10(2)165–199, https://econpapers.repec.org/repec:tsj:stataj:v:10:y:2010:i:2:p:165-199 (accessed 20 June 2021).
[2] Rao JN, Wu CF. Resampling inference with complex survey data. J Am Stat Assoc. 1998;83:401:231–241. DOI: 10.1080/01621459.1988.10478591.
[3] Rao JN, Wu CF, Yue K. Some recent work on resampling methods for complex surveys. Surv Methodol. 1992;18:209–217.

The variance is then estimated by examining the spread of the resulting estimates over the bootstrap samples. Let $\hat{p}_{(j)}$ denote the point estimate from the jth bootstrap sample and let R denote the total number of bootstrap samples (typically chosen to be 500). The estimate of the variance is given by the average of the squared deviations of the bootstrap estimates from the bootstrap mean:

$$\widehat{\mathrm{Var}}\left(\hat{p}^{\mathrm{ADR}}_{\mathrm{overall}}\right) = \frac{1}{R}\sum_{j=1}^{R}\left(\hat{p}_{(j)} - \bar{p}_{(.)}\right)^2$$

where $\bar{p}_{(.)}$ is the bootstrap mean, $\bar{p}_{(.)} = \frac{1}{R}\sum_{j=1}^{R}\hat{p}_{(j)}$. 95% confidence intervals can be constructed using a standard Wald formula (default in Stata) or using the quantiles of the bootstrap distribution.

Outcome 5: Prevalence of acquired HIV drug resistance to DTG among individuals with VNS and receiving ART who are taking a DTG-containing regimen.

Prevalence

The estimated prevalence of acquired DTG-specific drug resistance among individuals with VNS and receiving DTG-containing regimens follows a similar form as the overall estimate; however, it only uses data from individuals receiving DTG-containing regimens, it assesses for DTG-specific drug resistance, and the stratification weight, $\frac{N^{VNS}_{DTG}}{m^{VNS}_{DTG}}$, can be excluded because, given that it is constant, it does not affect the prevalence or variance estimates.

For each clinic k, the non-response sampling weight is defined as the number of sampled individuals receiving DTG-containing regimens with VNS, divided by the number of those who also have successful genotyping: $\frac{m^{VNS}_{DTG,k}}{m^{VNS,geno}_{DTG,k}}$. The non-response sampling weight is missing for individuals with VNS and missing genotype and is equal to 1 for individuals with viral suppression. The non-response weight assumes that genotyping failure is unrelated to the presence of ADR. Since the population is restricted to individuals with VNS, this is a subpopulation analysis and individuals with viral suppression or missing genotype will not contribute to the estimate but will contribute to the variance.

The final weight for patient l taking DTG-containing regimen in clinic k is given by $v_{DTG,k,l} = \frac{m^{VNS}_{DTG,k}}{m^{VNS,geno}_{DTG,k}}$. Let $X_{DTG,k,l}$ denote the binary outcome variable for acquired DTG-specific drug resistance among individuals with VNS and receiving DTG-containing regimens, equal to 1 if patient l in clinic k is resistant and 0 if not. The prevalence estimate is given by:

$$\hat{p}^{\mathrm{ADR}}_{DTG} = \frac{\sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m^{VNS,geno}_{DTG,k}} v_{DTG,k,l}X_{DTG,k,l}}{\sum_{k=1}^{c}\sum_{i\in I}\sum_{l=1}^{m^{VNS,geno}_{DTG,k}} v_{DTG,k,l}} = \frac{\sum_{k=1}^{c}\sum_{i\in I}\frac{m^{VNS}_{DTG,k}}{m^{VNS,geno}_{DTG,k}}\sum_{l=1}^{m^{VNS,geno}_{DTG,k}} X_{DTG,k,l}}{\sum_{k=1}^{c}\sum_{i\in I} m^{VNS}_{DTG,k}}.$$

Variance

Given the small sample sizes available for estimation of acquired DTG-specific drug resistance, a cluster bootstrap method rather than linearization is recommended for variance estimation. The bootstrap method involves generating a large number of independent resamples (that is, bootstrap samples) each drawn with replacement from the original sample. To address the clustering effect of ADR within clinics, the method resamples $c - 1$ clinics rather than individual specimens. For each of the bootstrap samples, the prevalence estimate specified above is calculated. The variance is estimated by examining the spread of the resulting estimates over the bootstrap samples.

To obtain a bootstrap variance estimator that is adjusted for bias, an internal rescaling procedure is applied to produce bootstrap replicate weights. The weight for replicate r is given by

$$v^{(r)}_{DTG,k,l} = \frac{c}{c-1} r_k v_{DTG,k,l}, r = 1, \ldots, R$$

where r_k is the number of times cluster k was resampled. These replicate weights represent the resamples and also allow variance estimates to be calculated with only knowledge of the first stage of sampling.

Let $\hat{p}_{(j)}$ denote the point estimate from the jth bootstrap sample and let R denote the total number of bootstrap samples (typically chosen to be 500). The estimate of the variance is given by the average of the squared deviations of the bootstrap estimates from the bootstrap mean:

$$\widehat{\mathrm{Var}}(\hat{p}_{DTG}^{ADR}) = \frac{1}{R}\sum_{j=1}^{R}\left(\hat{p}_{(j)} - \bar{p}_{(.)}\right)^2$$

where $\bar{p}_{(.)}$ is the bootstrap mean, $\bar{p}_{(.)} = \frac{1}{R}\sum_{j=1}^{R}\hat{p}_{(j)}$. 95% confidence intervals can be constructed using a standard Wald formula (default in Stata) or using the quantiles of the bootstrap distribution.

A2.4 Stata code

User-friendly instructions are provided below for data analysis in Stata (developed using Stata version 17.0). To use the code, data must follow the format described in sections 3.4–3.5, with patient-level and site-level information following the configuration of the Excel data upload template and HIV drug resistance sequences in FASTA file format. Sequence identification numbers and eligible case specimen identifiers must be identical and follow WHO convention.

In Stata, estimation and inference can be implemented using the svy package. The default variance estimation used is linearization (based on a first-order Taylor series linear approximation). Variance computations include finite population corrections when possible. Alternative statistical packages can be used to perform data analysis as long as they properly adjust for survey weights, clustering and stratification (if necessary). All statistical packages are expected to yield identical point estimates, but not all statistical packages are expected to yield identical standard error estimates and confidence intervals. Statistical packages that do not allow users to specify the finite population correction at each stage of sampling will overestimate the standard error, especially in countries with small eligible populations.

A2.4.1 Stata code for the survey in adults

The instructions below are for the survey in adults. Annex 2, sections A2.4.1.1–A2.4.1.7, provide Stata code for processing and combining the two data sets, and Annex 2, sections A2.4.1.8–A2.4.1.11, provide code for analysing the survey outcomes. All code displayed can be found in a downloadable Stata do-file that can run all pre-processing and analysis instructions at once, available at: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/treatment/hiv-drug-resistance/hiv-drug-resistance-surveillance/surveillance-of-acquired-hiv-drug-resistance-in-populations-receiving-art.

A2.4.1.1 Import viral load laboratory and patient-level data into Stata

Begin by importing viral load laboratory and patient-level data from the Excel file, named "patient_data_A3.xlsx" for this example. The Excel file should contain six sheets titled: (1) Survey Information - ADULT; (2) Configuration; (3) National Information - Adult; (4) ART Clinic Information - Adult; (5) Survey Participants - Adult; and (6) Participant Treatments - Adult.

1. To start, one may choose to create a do-file so that commands can be saved and then run. Click on the notepad icon corresponding to "New Do-file Editor" on the top-left corner of the Stata viewer, then save the do-file that is created.

2. Clear any previous output and set the working directory to the directory containing the data files. For example, if the directory is "C:/Documents", run the following code:

```
clear
cd "C:/Documents"
```

3. Import the "National Information - Adult" sheet of the Excel file, storing the first row as headings and changing all header names to uppercase. Save as a file named "NATIONAL_INFORMATION_ADULT.dta".

```
import excel using "patient_data_A3.xlsx", sheet("National Information - Adult") cellrange(A2) firstrow case(upper)
save "NATIONAL_INFORMATION_ADULT", replace
clear
```

4. Import the "ART Clinic Information - Adult" sheet of the Excel file, storing the first row as headings and changing all header names to uppercase. Save as a file named "ART_CLINIC_INFORMATION_ADULT.dta".

```
import excel using "patient_data_A3.xlsx", sheet("ART Clinic Information - Adult") cellrange(A2) firstrow ///
        case(upper)
save "ART_CLINIC_INFORMATION_ADULT", replace
clear
```

5. Import the "Survey Participants - Adult" sheet of the Excel file, storing the first row as headings and changing all header names to uppercase. Save as a file named "SURVEY_PARTICIPANTS_ADULT.dta".

```
import excel using "patient_data_A3.xlsx", sheet("Survey Participants - Adult") firstrow case(upper)
save "SURVEY_PARTICIPANTS_ADULT", replace
clear
```

6. Import the "Participant Treatments - Adult" sheet of the Excel file, storing the first row as headings and changing all header names to uppercase. Save as a file named "PARTICIPANT_TREATMENTS_ADULT.dta".

```
import excel using "patient_data_A3.xlsx", sheet("Participant Treatments - Adult") firstrow case(upper)
save "PARTICIPANT_TREATMENTS_ADULT", replace
clear
```

A2.4.1.2 Import HIV drug resistance data into Stata

1. The resistance data file should be an Excel file containing one sheet titled "ResistanceSummary". Import the HIV drug resistance data, storing the first row as headers and changing all header names to uppercase. Given the file name "FASTA_A3.xlsx", run:

```
import excel using "FASTA_A3.xlsx", sheet("ResistanceSummary") firstrow case(upper)
```

2. Rename SEQUENCENAME as PARTICIPANTID. Drop all cells without a subject ID. Drop all unnecessary variables. Replace "NA" and "None" as missing.

```
rename SEQUENCENAME PARTICIPANTID
drop if missing(PARTICIPANTID)
drop *SCORE ALGORITHM* STRAIN GENES PI* NRTI* NNRTI* INSTI*
destring, ignore("NA" "None") replace
```

3. For each of the resistance level variables, reclassify the variable as a binary resistance indicator, with levels 1–2 corresponding to susceptible (no HIV drug resistance) and levels 3–5 corresponding to HIV drug resistance. Rename resistance type variables.

```
ds *LEVEL
local plist = r(varlist)
foreach i of local plist {
        replace `i' = 0 if `i' < 3 & !missing(`i')
        replace `i' = 1 if `i' >= 3 & !missing(`i')
}
rename *LEVEL *_RES
```

4. Generate variables of DTG-specific resistance, any boosted PI resistance, any NRTI resistance, any NNRTI resistance, any INI resistance and any acquired HIV drug resistance, according to the definitions in section 3.5.

```
gen DTG_ADR = DTG_RES
egen ANY_PI = rowmax(ATVR_RES DRVR_RES LPVR_RES)
egen ANY_NRTI = rowmax(ABC_RES AZT_RES D4T_R DDI_RES FTC_RES TDF_RES)
egen ANY_NNRTI = rowmax(EFV_RES NVP_RES)
egen ANY_INI = rowmax(BIC_RES DTG_RES EVG_RES RAL_RES)
egen ANY_ADR = rowmax(ANY_PI ANY_NRTI ANY_NNRTI ANY_INI)
```

5. Save the modified HIV drug resistance data as a .dta file. In this example, we save the data as "RESISTANCE_SUMMARY.dta".

```
save RESISTANCE_SUMMARY, replace
```

A2.4.1.3. Prepare national-level data

1. Remove the previous data set, then load the national-level data, stored as "NATIONAL_INFORMATION_ADULT.dta". Rename the variables for total number of clinics providing ART to adults, total number of adults receiving ART, number of clinics serving adults to be sampled and proportion of adults receiving DTG-containing regimens.

```
clear
use NATIONAL_INFORMATION_ADULT.dta
rename NADULTCLINICSSAMPLINGTABLE C_ADULT
rename NADULTSARTSAMPLINGTABLE N_ADULT
rename NADULTCLINICSSAMPLED SAMPLE_C_ADULT
rename PROPADULTSONDTG P_DTG_ADULT
```

2. Save the modified data as a .dta file. In this example, we save the data as "NATIONAL_INFORMATION_ADULT.dta".

```
save NATIONAL_INFORMATION_ADULT, replace
```

A2.4.1.4. Prepare clinic-level data

3. Remove the previous data set, then load the clinic-level data, stored as "ART_CLINIC_INFORMATION_ADULT.dta", and rename the variables for site code and clinics size in the sampling table.

```
clear
use ART_CLINIC_INFORMATION_ADULT.dta
rename UNIQUE3LETTERCLINICCODE SITECODE
```

2.4.1.5 Prepare patient-level data on ARV drug treatment regimen

1. Remove the previous data set and then load the treatment regimen data, stored as "PARTICIPANT_TREATMENTS_ADULT.dta".

```
clear
use PARTICIPANT_TREATMENTS_ADULT.dta
```

2. Exclude observations missing a subject ID or belonging to the child/adolescent population. Rename ARV drug types so that all variable names begin with a letter.

```
drop if missing(PARTICIPANTID) | substr(PARTICIPANTID, 23, 1) != "a"
replace ARVDRUG = "ARV_" + ARVDRUG
```

3. Rename prior ARV drug types to differentiate them from current drug types and then drop the current ART indicator variable.

```
replace ARVDRUG = "PRIOR_" + ARVDRUG if CURRENTARTYN == "N"
drop CURRENTARTYN
```

4. Generate an indicator variable of DTG-containing ART, equal to 1 if a person is taking a DTG-containing regimen and 0 if a person is taking a non-DTG-containing regimen.

```
gen TEMP_DTG = cond(inlist(ARVDRUG, "ARV_DTG", "ARV_TLD", "ARV_JUL"), 1, 0)
by PARTICIPANTID, sort: egen DTG = max(TEMP_DTG)
drop TEMP_DTG
```

5. Reformat the ARVDRUG variable so that each ARV drug type is created as a new binary variable. Set to 1 if the person's regimen includes the drug and 0 if not.

```
gen ON = 1
reshape wide ON, i(PARTICIPANTID) j(ARVDRUG) string
rename ON* *
```

6. Save the modified data as a .dta file. In this example, we save the data as "PARTICIPANT_TREATMENTS_ADULT.dta".

```
save PARTICIPANT_TREATMENTS_ADULT, replace
```

2.4.1.6 Prepare patient-level data on other variables

1. Remove the previous dataset and then load the patient-level data, stored as "SURVEY_PARTICIPANTS_ADULT.dta".

```
clear
use SURVEY_PARTICIPANTS_ADULT.dta
```

2. Recode unknown values as missing.

```
recode DATE* (9999 = .)
recode AGE (-9 = .)
foreach var of varlist BREASTFEEDINGSTATUS PREGNANCYSTATUS CURRENTART {
        replace `var' = "." if `var' == "UNK"
}
```

3. Drop observations with missing subject ID or belonging to the child/adolescent population. Drop lab specimen code variable.

```
drop if missing(PARTICIPANTID) | substr(PARTICIPANTID, 23, 1) != "a"
drop LABSPECIMENCODE
```

4. Convert the viral copies variable from string to integer and generate the viral suppression variable (<1000 copies/mL).

```
destring VIRALLOADVALUECOPIESML, replace
gen VIRAL_SUPPRESSION = cond(VIRALLOADVALUECOPIESML < 1000, 1, 0) if ///
        VIRALLOADVALUECOPIESML != .
```

5. Save the modified data as a .dta file. In this example, we save the data as "SURVEY_PARTICIPANTS_ADULT.dta".

```
save SURVEY_PARTICIPANTS_ADULT.dta, replace
```

A2.4.1.7 Merge all data sets

1. Merge in the national data, dropping variables and observations that are unnecessary.

```
clear
use SURVEY_PARTICIPANTS_ADULT.dta
gen N_ADULT = .
append using NATIONAL_INFORMATION_ADULT, keep(N_ADULT C_ADULT SAMPLE_C_ADULT ///
        P_DTG_ADULT)
drop if missing(PARTICIPANTID) & missing(N_ADULT)
foreach var of varlist N_ADULT C_ADULT SAMPLE_C_ADULT P_DTG_ADULT {
        replace `var' = `var'[_N]
}
drop if missing(PARTICIPANTID) | substr(PARTICIPANTID, 23, 1) != "a"
```

2. Use a many-to-one merge to merge the clinic-level data on the site code.

```
merge m:1 SITECODE using ART_CLINIC_INFORMATION_ADULT, keep(match) nogenerate
```

3. Merge in the treatment regimen data by subject ID.

```
merge 1:1 PARTICIPANTID using PARTICIPANT_TREATMENTS_ADULT, keep(match) nogenerate
```

4. Merge in the HIV drug resistance data by subject ID.

```
merge 1:1 PARTICIPANTID using RESISTANCE_SUMMARY, keep(match) nogenerate
```

5. Save the combined and reorganized data as a .dta file. In this example, we save the data as "ALL_DATA.dta".

```
save ALL_DATA.dta, replace
```

A2.4.1.8. Create survey weights and other necessary variables and declare survey design

1. Remove the previous dataset and load in the combined data.
```
clear
use ALL_DATA.dta
```

2. Generate the first stage site sampling weight, calculated as the total population, divided by the product of the number of clinics to be sampled and the clinic sizes.
```
by SITECODE, sort: gen WEIGHT1 = N_ADULT/(SAMPLE_C_ADULT*CLINIC_SIZE)
```

3. Create the variable for stratification by regimen (DTG versus non-DTG).
```
egen STRATA = group(DTG)
```

4. Generate the variable for total DTG and non-DTG case specimens per sampled clinic. This corresponds to the second-stage populations.
```
gen CLINIC_POP = cond(DTG == 1, N_ADULT_DTG, N_ADULT_NONDTG)
```

5. Generate the second stage sampling weight, calculated as the number of adults receiving ART observed during the survey period, divided by the number of specimens sampled and with successful viral load testing.
```
by SITECODE DTG, sort: gen WEIGHT2 = CLINIC_POP/_N
```

6. Generate the sampling weight as the product of the first- and second-stage sampling weights.
```
gen WEIGHTS = WEIGHT1 * WEIGHT2
```

7. Set the stratified two-stage clustered survey design with finite population correction. If there exists a stratum or multiple strata with only one unit sampled, sampling errors cannot be estimated for all strata independently, and Stata will report a missing standard error. We recommend addressing this by setting the standard errors for single-unit strata to be the average of the standard errors for other strata. This is represented by the single unit (scaled) term in the code below.
```
svyset SITECODE [pweight = WEIGHTS], fpc(C_ADULT) || _n, strata(STRATA) fpc(CLINIC_POP) ///
            singleunit(scaled)
```

A2.4.1.9 Analysis for outcomes 1 and 2

1. Obtain estimates and confidence intervals for the prevalence of viral suppression among all adults receiving ART. In the output, the point estimate, standard error and 95% confidence interval of interest are located in the row labelled "1".
```
svy: proportion VIRAL_SUPPRESSION
```
This command gives confidence intervals expressed on the logit scale. For Wald confidence intervals, simply add citype(wald)to the end:
```
svy: proportion VIRAL_SUPPRESSION, citype(wald)
```
The design effect can also be obtained:
```
estat effects
```

2. Obtain estimates and confidence intervals for the prevalence of overall drug resistance among adults receiving DTG regimens.
```
svy, subpop(if DTG==1): proportion VIRAL_SUPPRESSION
```

3. Store the estimated prevalence of viral suppression among adults receiving DTG-containing regimens for use in ADR estimates.
```
gen P_DTG_VS = e(b)[1,2]
```

4. Obtain estimates and confidence intervals for the prevalence of overall drug resistance among adults receiving non-DTG regimens.
```
svy, subpop(if DTG==0): proportion VIRAL_SUPPRESSION
```

5. Store the estimated prevalence of viral suppression among adults receiving non-DTG-containing regimens for use in ADR estimates.
```
gen P_NONDTG_VS = e(b)[1,2]
```

A2.4.1.10 Analysis for outcome 3

Some examples of subgroup analysis are given below.

1. Obtain prevalence and variance estimates of viral suppresion among men.

```
svy, subpop(if GENDER == "M"): proportion VIRAL_SUPPRESSION
```

2. Obtain prevalence and variance estimates of viral suppression among adults who are breastfeeding.

```
svy, subpop(if BREASTFEEDINGSTATUS == "Y"): proportion VIRAL_SUPPRESSION
```

A2.4.1.11 Analysis for outcomes 4 and 5

1. Install the "bsweights" package to run bootstrap variance estimation in complex survey data. Generate a byte variable as a placeholder for the strata variable.

```
net install bsweights, from(http://staskolenikov.net/stata)
generate byte _one=1
```

2. Generate the weight that accounts for stratification by regimen. Obtain the regimen-specific total number of individuals observed during the survey period, sampled across clinics. Then, multiply by the estimated regimen-specific VNS prevalence using Outcomes 1 and 2 to obtain the total with VNS. Obtain the regimen-specific total number of sampled individuals with VNS. Generate the stratification weight by dividing the regimen-specific total with VNS by the regimen-specific number of sampled individuals with VNS.

```
by DTG, sort: egen REG_TOTAL = total(CLINIC_POP)
gen REG_TOTAL_VNS = REG_TOTAL*cond(DTG==1, 1-P_DTG_VS, 1-P_NONDTG_VS)
by DTG, sort: egen REG_SAMPLED_TOTAL_VNS = total(VIRAL_SUPPRESSION == 0) if ///
        VIRAL_SUPPRESSION == 0
gen STRAT_WT = REG_TOTAL_VNS / REG_SAMPLED_TOTAL_VNS
```

3. Generate the weight that accounts for non-response due to genotyping failure. Obtain the total number of sampled individuals with VNS, for each regimen and each clinic. Then, obtain the total number of sampled individuals with VNS and successful genotyping for each regimen and each clinic. The non-response weight is the regimen-specific number sampled with viral non-suppression divided by the regimen-specific number sampled with VNS and successful genotyping. The non-response weight is 1 for individuals with viral suppression.

```
by SITECODE DTG, sort: egen REG_SAMPLED_VNS = total(VIRAL_SUPPRESSION == 0) if ///
        VIRAL_SUPPRESSION == 0
by SITECODE DTG, sort: egen REG_SAMPLED_VNS_GENO = total(ANY_ADR != .) if ANY_ADR != .
gen NON_RESPONSE_WT = REG_SAMPLED_VNS / REG_SAMPLED_VNS_GENO
replace NON_RESPONSE_WT = 1 if VIRAL_SUPPRESSION == 1
```

4. Generate the sampling weight as the product of the stratification and non-response weights. Clinic sampling weights are ignored.

```
gen ADR_WT = STRAT_WT*NON_RESPONSE_WT
```

5. Set the initial survey design with default linearized variance. Obtain bootstrap replication weights with 100 replications and c-1 resampled units. Then, define the new survey design with bootstrap replication weights and the cluster bootstrap variance. Note that the finite population correction cannot be used with the cluster bootstrap.

```
svyset SITECODE [pw=ADR_WT], strata(_one)|| _n, strata(STRATA) fpc(CLINIC_POP) singleunit(scaled)
bsweights bsw, reps(100) n(-1)
svyset SITECODE [pweight=ADR_WT], strata(_one) bsrweight(bsw*) vce(bootstrap) || _n, strata(STRATA) ///
        singleunit(scaled)
```

6. Obtain estimates and confidence intervals for the prevalence of overall drug resistance among all adults with VNS.

```
svy, subpop(if VIRAL_SUPPRESSION==0): proportion ANY_ADR
```

The design effect can also be obtained:

```
estat effects
```

Obtain estimates and confidence intervals for the prevalence of DTG-specific drug resistance among individuals taking DTG-containing regimens and with VNS.

```
svy, subpop(if DTG==1 & VIRAL_SUPPRESSION==0): proportion DTG_ADR
```

A2.4.2 Stata code for the survey among children and adolescents

Stata code for separate survey among children and adolescents

If the survey among children and adolescents is implemented separately from the survey among adults, the Stata code will be very similar to that of the survey among adults. The variable names may differ, but the structure is the same. A downloadable Stata do-file includes pre-processing and analysis instructions for this survey. The Stata do-file is available at: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/treatment/hiv-drug-resistance/hiv-drug-resistance-surveillance/surveillance-of-acquired-hiv-drug-resistance-in-populations-receiving-art.

Stata code for combined survey among children and adolescents

The below instructions are displayed for the combined survey among children and adolescents, which carries over clinics serving both children and adolescents that were selected for the survey among adults. Annex 2, sections A2.4.2.1–A2.4.2.7, provides Stata code for processing and combining the two data sets, and Annex 2, sections A2.4.2.8–A2.4.2.11, provide code for analysing the survey outcomes. All code displayed can be found in a downloadable Stata do-file that can run all pre-processing and analysis instructions at once. The Stata-do file is available at: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/treatment/hiv-drug-resistance/hiv-drug-resistance-surveillance/surveillance-of-acquired-hiv-drug-resistance-in-populations-receiving-art.

A2.4.2.1. Import viral load laboratory and patient-level data into Stata

Begin by importing viral load laboratory and patient-level data from the Excel upload template file, named "patient_data_A3_child.xlsx" for this example. The Excel file should contain six sheets titled: (1) Survey Information-CHILDREN+ADO, (2) Configuration, (3) National Information, (4) ART Clinic Information, (5) Survey Participants, and (6) Participant Treatments-ChildAdo.

1. To start, one may choose to create a do-file so that commands can be saved and then run. Click on the notepad icon corresponding to "New Do-file Editor" on the top-left corner of the Stata viewer, then save the do-file that is created.

2. Clear any previous output and set the working directory to the directory containing the data files. For example, if the directory is "C:/Documents", run the following code:

```
clear
cd "C:/Documents"
```

3. Import the "National Information" sheet of the Excel file, storing the first row as headings and changing all header names to uppercase. Save as a file named "NATIONAL_INFORMATION_CHILD_ADO.dta".

```
import excel using "patient_data_A3_child.xlsx", sheet("National Information") cellrange(A2) firstrow case(upper)
save "NATIONAL_INFORMATION_CHILD_ADO", replace
clear
```

4. Import the "ART Clinic Information" sheet of the Excel file, storing the first row as headings and changing all header names to uppercase. Save as a file named "ART_CLINIC_INFORMATION_CHILD_ADO.dta".

```
import excel using "patient_data_A3_child.xlsx", sheet("ART Clinic Information") cellrange(A2) firstrow case(upper)
save "ART_CLINIC_INFORMATION_CHILD_ADO", replace
clear
```

5. Import the "Survey Participants" sheet of the Excel file, storing the first row as headings and changing all header names to uppercase. Save as a file named "SURVEY_PARTICIPANTS_CHILD_ADO.dta".

```
import excel using "patient_data_A3_child.xlsx", sheet("Survey Participants") firstrow case(upper)
save "SURVEY_PARTICIPANTS_CHILD_ADO", replace
clear
```

6. Import the "Participant Treatments-ChildAdo" sheet of the Excel file, storing the first row as headings and changing all header names to uppercase. Save as a file named "PARTICIPANT_TREATMENTS_CHILD_ADO.dta".

```
import excel using "patient_data_A3_child.xlsx", sheet("Participant Treatments-ChildAdo") firstrow case(upper)
save "PARTICIPANT_TREATMENTS_CHILD_ADO", replace
clear
```

A2.4.2.2 Import HIV drug resistance data into Stata

1. The resistance data file should be an Excel file containing one sheet entitled "ResistanceSummary". Import the HIV drug resistance data, storing the first row as headers and changing all header names to uppercase. Given the file name "FASTA_A3_child.xlsx", run:

```
import excel using "FASTA_A3_child.xlsx", sheet("ResistanceSummary") firstrow case(upper)
```

2. Rename SEQUENCENAME as PARTICIPANTID. Drop all cells without a subject ID. Drop all unnecessary variables. Replace "NA" and "None" as missing.

```
rename SEQUENCENAME PARTICIPANTID
drop if missing(PARTICIPANTID)
drop *SCORE ALGORITHM* STRAIN GENES PI* NRTI* NNRTI* INSTI*
destring, ignore("NA" "None") replace
```

3. For each of the resistance level variables, reclassify the variable as a binary resistance indicator, with levels 1–2 corresponding to susceptible (no HIV drug resistance), and levels 3–5 corresponding to HIV drug resistance. Rename resistance type variables.

```
ds *LEVEL
local plist = r(varlist)
foreach i of local plist {
        replace `i' = 0 if `i' < 3 & !missing(`i')
        replace `i' = 1 if `i' >= 3 & !missing(`i')
}
rename *LEVEL *_RES
```

4. Generate variables of DTG-specific resistance, any boosted PI resistance, any NRTI resistance, any NNRTI resistance, any INI resistance and any acquired HIV drug resistance, according to the definitions in section 3.5.

```
gen DTG_ADR = DTG_RES
egen ANY_PI = rowmax(ATVR_RES DRVR_RES LPVR_RES)
egen ANY_NRTI = rowmax(ABC_RES AZT_RES D4T_R DDI_RES FTC_RES TDF_RES)
egen ANY_NNRTI = rowmax(EFV_RES NVP_RES)
egen ANY_INI = rowmax(BIC_RES DTG_RES EVG_RES RAL_RES)
egen ANY_ADR = rowmax(ANY_PI ANY_NRTI ANY_NNRTI ANY_INI)
```

5. Save the modified HIV drug resistance data as a .dta file. In this example, we save the data as "RESISTANCE_SUMMARY.dta".

```
save RESISTANCE_SUMMARY, replace
```

A2.4.2.3 Prepare national-level data

1. Remove the previous data set, then load the national-level data, stored as "NATIONAL_INFORMATION_CHILD_ADO.dta". Rename the variables for total number of clinics providing ART to adults, total number of adults receiving ART, number of clinics serving adults to be sampled, and proportion of adults receiving DTG-containing regimens.

```
clear
use NATIONAL_INFORMATION_CHILD_ADO.dta
rename NADULTCLINICSSAMPLINGTABLE C_ADULT
rename NADULTSARTSAMPLINGTABLE N_ADULT
rename NADULTCLINICSSAMPLED SAMPLE_C_ADULT
rename PROPADULTSONDTG P_DTG_ADULT
```

2. Rename the variables for: total number of clinics providing ART to children/adolescents, excluding clinics selected for the adult survey; total number of children/adolescents receiving ART, excluding children/adolescents selected for the adult survey; number of clinics serving children/adolescents to be sampled; and proportion of children and adolescents receiving DTG-containing regimens.

```
rename NCHILDADOCLINICSSAMPTABLE C_CHILD
rename NCHILDADOARTSAMPLINGTABLE N_CHILD
rename NCHILDADOCLINICSSAMPLED SAMPLE_C_CHILD
rename PROPCHILDADOONDTG P_DTG_CHILD
```

3. Save the modified data as a .dta file. In this example, we save the data as "NATIONAL_INFORMATION_CHILD_ADO.dta".

```
save NATIONAL_INFORMATION_CHILD_ADO, replace
```

A2.4.2.4 Prepare clinic-level data

1. Remove the previous data set, then load the clinic-level data, stored as "ART_CLINIC_INFORMATION_CHILD_ADO.dta"

```
clear
use ART_CLINIC_INFORMATION_CHILD_ADO.dta
```

2. Rename the variables for site code, adult clinic size, child clinic size and clinics serving both adults and children/adolescents.

```
rename UNIQUE3LETTERCLINICCODE SITECODE
rename N_ADULTCLINIC_SIZE ADULT_CLINIC_SIZE
rename N_CHILDADOCLINIC_SIZE CHILD_CLINIC_SIZE
rename CLINIC_BOTHADULTCHILDADO_ART CLINIC_ADULT_CHILD_ART
```

3. Save the modified data as a .dta file. In this example, we save the data as "ART_CLINIC_INFORMATION_CHILD_ADO.dta".

```
save ART_CLINIC_INFORMATION_CHILD_ADO, replace
```

A2.4.2.5 Prepare patient-level data on ARV drug treatment regimen

1. Remove the previous data set and then load the treatment regimen data, stored as "PARTICIPANT_TREATMENTS_CHILD_ADO.dta".

```
clear
use PARTICIPANT_TREATMENTS_CHILD_ADO.dta
```

2. Exclude observations missing a subject ID or belonging to the adult population. Rename ARV drug types so that all variable names begin with a letter.

```
drop if missing(PARTICIPANTID) | substr(PARTICIPANTID, 23, 1) != "c"
replace ARVDRUG = "ARV_" + ARVDRUG
```

3. Rename prior ARV drug types to differentiate them from current drug types and then drop the current ART indicator variable.

```
replace ARVDRUG = "PRIOR_" + ARVDRUG if CURRENTARTYN == "N"
drop CURRENTARTYN
```

4. Generate an indicator variable of DTG-containing ART, equal to 1 if a person is taking a DTG-containing regimen and 0 if a person is taking a non-DTG-containing regimen.

```
gen TEMP_DTG = cond(inlist(ARVDRUG, "ARV_DTG", "ARV_TLD", "ARV_JUL"), 1, 0)
by PARTICIPANTID, sort: egen DTG = max(TEMP_DTG)
drop TEMP_DTG
```

5. Reformat the ARVDRUG variable so that each ARV drug type is created as a new binary variable. Set to 1 if the person's regimen includes the drug and 0 if not.

```
gen ON = 1
reshape wide ON, i(PARTICIPANTID) j(ARVDRUG) string
rename ON* *
```

6. Save the modified data as a .dta file. In this example, we save the data as "PARTICIPANT_TREATMENTS_CHILD_ADO.dta".

```
save PARTICIPANT_TREATMENTS_CHILD_ADO, replace
```

A2.4.2.6 Prepare patient-level data on other variables

1. Remove the previous dataset and then load the patient-level data, stored as "SURVEY_PARTICIPANTS_CHILD_ADO.dta".

```
clear
use SURVEY_PARTICIPANTS_CHILD_ADO.dta
```

2. Recode unknown values as missing.

```
recode DATE* (9999 = .)
recode AGE (-9 = .)
foreach var of varlist BREASTFEEDINGSTATUS PREGNANCYSTATUS CURRENTART {
        replace `var' = "." if `var' == "UNK"
}
```

3. Drop observations with missing subject ID or belonging to the adult population. Drop lab specimen code variable.

```
drop if missing(PARTICIPANTID) | substr(PARTICIPANTID, 23, 1) != "c"
drop LABSPECIMENCODE
```

4. Convert the viral copies variable from string to integer and generate the viral suppression variable (<1000 copies/mL).

```
destring VIRALLOADVALUECOPIESML, replace
gen VIRAL_SUPPRESSION = cond(VIRALLOADVALUECOPIESML < 1000, 1, 0) if ///
        VIRALLOADVALUECOPIESML != .
```

5. Save the modified data as a .dta file. In this example, we save the data as "SURVEY_PARTICIPANTS_CHILD_ADO.dta".

```
save SURVEY_PARTICIPANTS_CHILD_ADO.dta, replace
```

A2.4.2.7 Merge all data sets

1. Merge in the national data, dropping variables and observations that are unnecessary.

```
clear
use SURVEY_PARTICIPANTS_CHILD_ADO.dta
gen N_ADULT = .
append using NATIONAL_INFORMATION_CHILD_ADO, keep(N_ADULT C_ADULT SAMPLE_C_ADULT ///
        P_DTG_ADULT N_CHILD C_CHILD SAMPLE_C_CHILD P_DTG_CHILD)
drop if missing(PARTICIPANTID) & missing(N_ADULT)
foreach var of varlist N_ADULT C_ADULT SAMPLE_C_ADULT P_DTG_ADULT N_CHILD C_CHILD ///
        SAMPLE_C_CHILD P_DTG_CHILD {
                replace `var' = `var'[_N]
}
drop if missing(PARTICIPANTID) | substr(PARTICIPANTID, 23, 1) != "c"
```

2. Use a many-to-one merge to merge the clinic-level data on the site code.

```
merge m:1 SITECODE using ART_CLINIC_INFORMATION_CHILD_ADO, keep(match) nogenerate
```

3. Merge in the treatment regimen data by subject ID.

```
merge 1:1 PARTICIPANTID using PARTICIPANT_TREATMENTS_CHILD_ADO, keep(match) nogenerate
```

4. Merge in the HIV drug resistance data by subject ID.

```
merge 1:1 PARTICIPANTID using RESISTANCE_SUMMARY, keep(match) nogenerate
```

5. Save the combined and reorganized data as a .dta file. In this example, we save the data as "ALL_DATA.dta".

```
save ALL_DATA.dta, replace
```

A2.4.2.8 Create survey weights and other necessary variables and declare survey design

1. Remove the previous dataset and load in the combined data.

```
clear
use ALL_DATA.dta
```

2. Generate the first stage site sampling weight, calculated as described in Annex 2, section A2.3.2.2, with the form depending on whether the clinic serves both adults and children and adolescents or only children and adolescents.

2. Generate the first stage site sampling weight, calculated as described in Annex 2, section A2.3.2.2, with the form depending on whether the clinic serves both adults and children and adolescents or only children and adolescents.

```
by SITECODE, sort: gen P_SELECT_ADULT_CHILD = (SAMPLE_C_ADULT*ADULT_CLINIC_SIZE) /  ///
            N_ADULT + ((SAMPLE_C_CHILD*CHILD_CLINIC_SIZE)/N_CHILD) *  ///
            (1-((SAMPLE_C_ADULT*ADULT_CLINIC_SIZE)/N_ADULT))  ///
            if CLINIC_ADULT_CHILD_ART == "Y"
by SITECODE, sort: gen P_SELECT_CHILD_ONLY = (SAMPLE_C_CHILD*CHILD_CLINIC_SIZE)/N_CHILD ///
            if CLINIC_ADULT_CHILD_ART == "N"
by SITECODE, sort: gen WEIGHT1 = 1/P_SELECT_ADULT_CHILD if CLINIC_ADULT_CHILD_ART == "Y"
by SITECODE, sort: replace WEIGHT1 = 1/P_SELECT_CHILD_ONLY if CLINIC_ADULT_CHILD_ART == "N"
```

3. Create the variable for stratification by regimen (DTG versus non-DTG).

```
egen STRATA = group(DTG)
```

4. Generate the variable for total DTG and non-DTG case specimens per sampled clinic. This corresponds to the second-stage populations.

```
gen CLINIC_POP = cond(DTG == 1, N_CHILD_ADO_DTG, N_CHILD_ADO_NONDTG)
```

5. Generate the second-stage sampling weight, calculated as the number of children and adolescents receiving ART observed during the survey period, divided by the number of specimens sampled and with successful viral load testing.

```
by SITECODE DTG, sort: gen WEIGHT2 = CLINIC_POP/_N
```

6. Generate the sampling weight as the product of the first- and second-stage sampling weights.

```
gen WEIGHTS = WEIGHT1 * WEIGHT2
```

7. Set the stratified two-stage clustered survey design with finite population correction. If there exists a stratum or multiple strata with only one unit sampled, sampling errors cannot be estimated for all strata independently, and Stata will report a missing standard error. We recommend addressing this by setting the standard errors for single-unit strata to be the average of the standard errors for other strata. This is represented by the single unit (scaled) term in the code below.

```
svyset SITECODE [pweight = WEIGHTS], fpc(C_CHILD) || _n, strata(STRATA) fpc(CLINIC_POP) ///
            singleunit(scaled)
```

A2.4.2.9 Analysis for outcomes 1 and 2

1. Obtain estimates and confidence intervals for the prevalence of viral suppression among all children/adolescents receiving ART. In the output, the point estimate, standard error and 95% confidence interval of interest are located in the row labelled "1".

```
svy: proportion VIRAL_SUPPRESSION
```

This command gives confidence intervals expressed on the logit scale. For Wald confidence intervals, simply add citype(wald)to the end:

```
svy: proportion VIRAL_SUPPRESSION, citype(wald)
```

The design effect can also be obtained:

```
estat effects
```

2. Obtain estimates and confidence intervals for the prevalence of overall drug resistance among children/adolescents receiving DTG-containing regimens.

```
svy, subpop(if DTG==1): proportion VIRAL_SUPPRESSION
```

3. Store the estimated prevalence of viral suppression among children/adolescents receiving DTG-containing regimens for use in ADR estimates.

```
gen P_DTG_VS = e(b)[1,2]
```

4. Obtain estimates and confidence intervals for the prevalence of overall drug resistance among children/adolescents receiving non-DTG-containing regimens.

```
svy, subpop(if DTG==0): proportion VIRAL_SUPPRESSION
```

5. Store the estimated prevalence of viral suppression among children/adolescents receiving non-DTG-containing regimens for use in ADR estimates.

```
gen P_NONDTG_VS = e(b)[1,2]
```

A2.4.2.10. Analysis for outcome 3

Some examples of subgroup analysis are given below.

3. Obtain prevalence and variance estimates of viral suppression among child/adolescent men.

```
svy, subpop(if GENDER == "M"): proportion VIRAL_SUPPRESSION
```

4. Obtain prevalence and variance estimates of viral suppression among children/adolescents who are breastfeeding.

```
svy, subpop(if BREASTFEEDINGSTATUS == "Y"): proportion VIRAL_SUPPRESSION
```

A2.4.2.11 Analysis for outcomes 4 and 5

1. Install the "bsweights" package to run bootstrap variance estimation in complex survey data. Generate a byte variable as a placeholder for the strata variable.

```
net install bsweights, from(http://staskolenikov.net/stata)
generate byte _one=1
```

2. Generate the weight that accounts for stratification by regimen. Obtain the regimen-specific total number of individuals observed during the survey period, sampled across clinics. Then, multiply by the estimated regimen-specific VNS prevalence using Outcomes 1 and 2 to obtain the total with VNS. Obtain the regimen-specific total number of sampled individuals with VNS. Generate the stratification weight by dividing the regimen-specific total with VNS by the regimen-specific number of sampled individuals with VNS.

```
by DTG, sort: egen REG_TOTAL = total(CLINIC_POP)
gen REG_TOTAL_VNS = REG_TOTAL*cond(DTG==1, 1-P_DTG_VS, 1-P_NONDTG_VS)
by DTG, sort: egen REG_SAMPLED_TOTAL_VNS = total(VIRAL_SUPPRESSION == 0) if ///
        VIRAL_SUPPRESSION == 0
gen STRAT_WT = REG_TOTAL_VNS / REG_SAMPLED_TOTAL_VNS
```

3. Generate the weight that accounts for non-response due to genotyping failure. Obtain the total number of sampled individuals with VNS, for each regimen and each clinic. Then, obtain the total number of sampled individuals with VNS and successful genotyping for each regimen and each clinic. The non-response weight is the regimen-specific number sampled with VNS divided by the regimen-specific number sampled with VNS and successful genotyping. The non-response weight is 1 for individuals with viral suppression.

```
by SITECODE DTG, sort: egen REG_SAMPLED_VNS = total(VIRAL_SUPPRESSION == 0) if ///
        VIRAL_SUPPRESSION == 0
by SITECODE DTG, sort: egen REG_SAMPLED_VNS_GENO = total(ANY_ADR != .) if ANY_ADR != .
gen NON_RESPONSE_WT = REG_SAMPLED_VNS / REG_SAMPLED_VNS_GENO
replace NON_RESPONSE_WT = 1 if VIRAL_SUPPRESSION == 1
```

4. Generate the sampling weight as the product of the stratification and non-response weights. Clinic sampling weights are ignored.

```
gen ADR_WT = STRAT_WT*NON_RESPONSE_WT
```

5. Set the initial survey design with default linearized variance. Obtain bootstrap replication weights with 100 replications and c-1 resampled units. Then, define the new survey design with bootstrap replication weights and the cluster bootstrap variance. Note that the finite population correction cannot be used with the cluster bootstrap.

```
svyset SITECODE [pw=ADR_WT], strata(_one)|| _n, strata(STRATA) fpc(CLINIC_POP) singleunit(scaled)
bsweights bsw, reps(100) n(-1)
svyset SITECODE [pweight=ADR_WT], strata(_one) bsrweight(bsw*) vce(bootstrap) || _n, strata(STRATA) ///
        singleunit(scaled)
```

6. Obtain estimates and confidence intervals for the prevalence of overall drug resistance among all children/adolescents with VNS.

```
svy, subpop(if VIRAL_SUPPRESSION==0): proportion ANY_ADR
```

The design effect can also be obtained:

```
estat effects
```

7. Obtain estimates and confidence intervals for the prevalence of DTG-specific drug resistance among children/adolescents taking DTG-containing regimens and with VNS.

```
svy, subpop(if DTG==1 & VIRAL_SUPPRESSION==0): proportion DTG_ADR
```

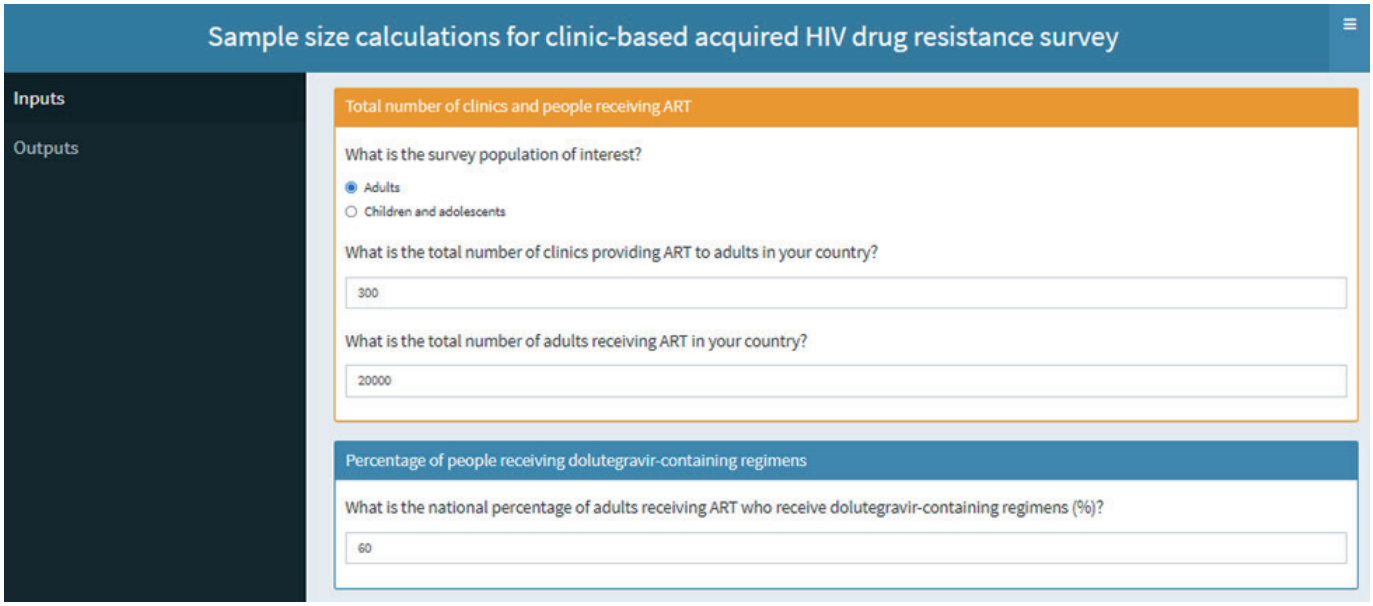
ANNEX 3. EXAMPLE: SAMPLE SIZE CALCULATIONS USING THE METHODS DESCRIBED IN SECTION 2.4

This annex provides an example of how to calculate country-specific per-clinic and total sample sizes using the WHO online sample size calculator (available at https://worldhealthorg.shinyapps.io/ADR_ClinicBasedMethod/).

Inputs

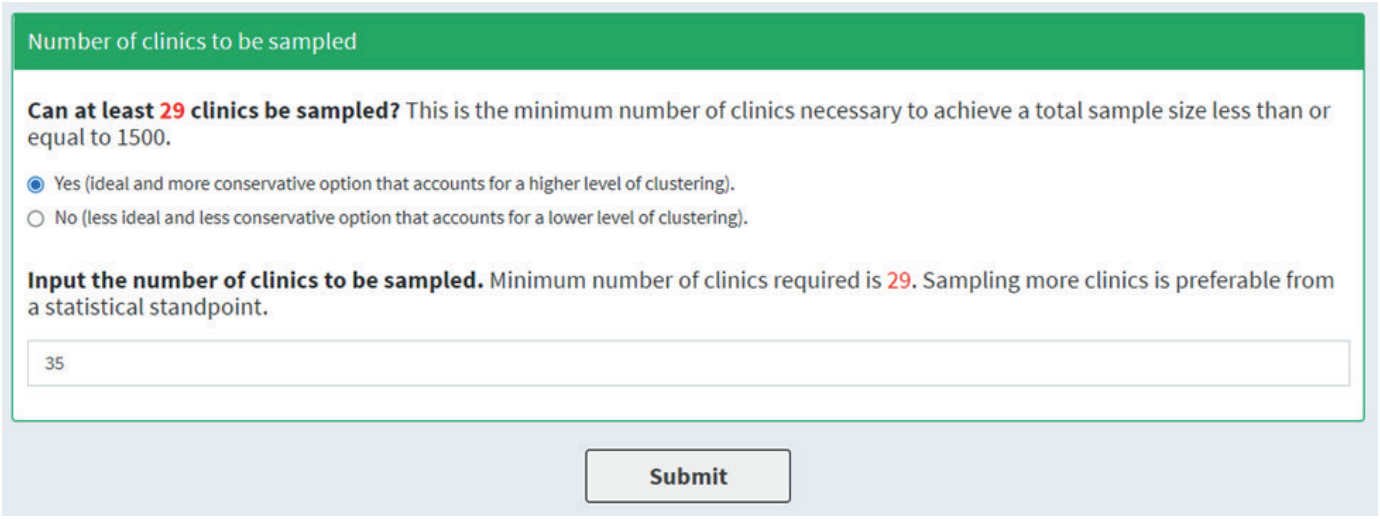
Suppose a country has 300 clinics supporting adults and 20 000 total adults receiving ART, and 60% of adults receiving ART are on DTG-containing regimens. In the online sample size calculator, the user can begin by specifying these three values. (See Fig. A3.1).

Fig. A3.1. Online sample size calculator: example input number of clinics, number of people receiving ART and percentage receiving DTG-containing regimens



Next, the application will display the minimum number of clinics necessary to ensure valid per clinic sample sizes and a total sample size that is less than or equal to 1500 specimens (see Fig. A3.2). The user can choose whether or not this number of clinics can be sampled. A choice of "yes" will result in a sample size calculation that accounts for a higher level of clustering and is more conservative. A choice of "no" will account for a lower level of clustering and is, therefore, less ideal but may aid feasibility. Choosing "no" will also trigger a display for a smaller minimum number of clinics. After choosing "yes" or "no," the user can then input the actual number of clinics that are to be sampled, and then click "Submit." For this example, a minimum number of 29 clinics is necessary. Suppose the country chooses to sample 35 clinics to improve precision of estimates.

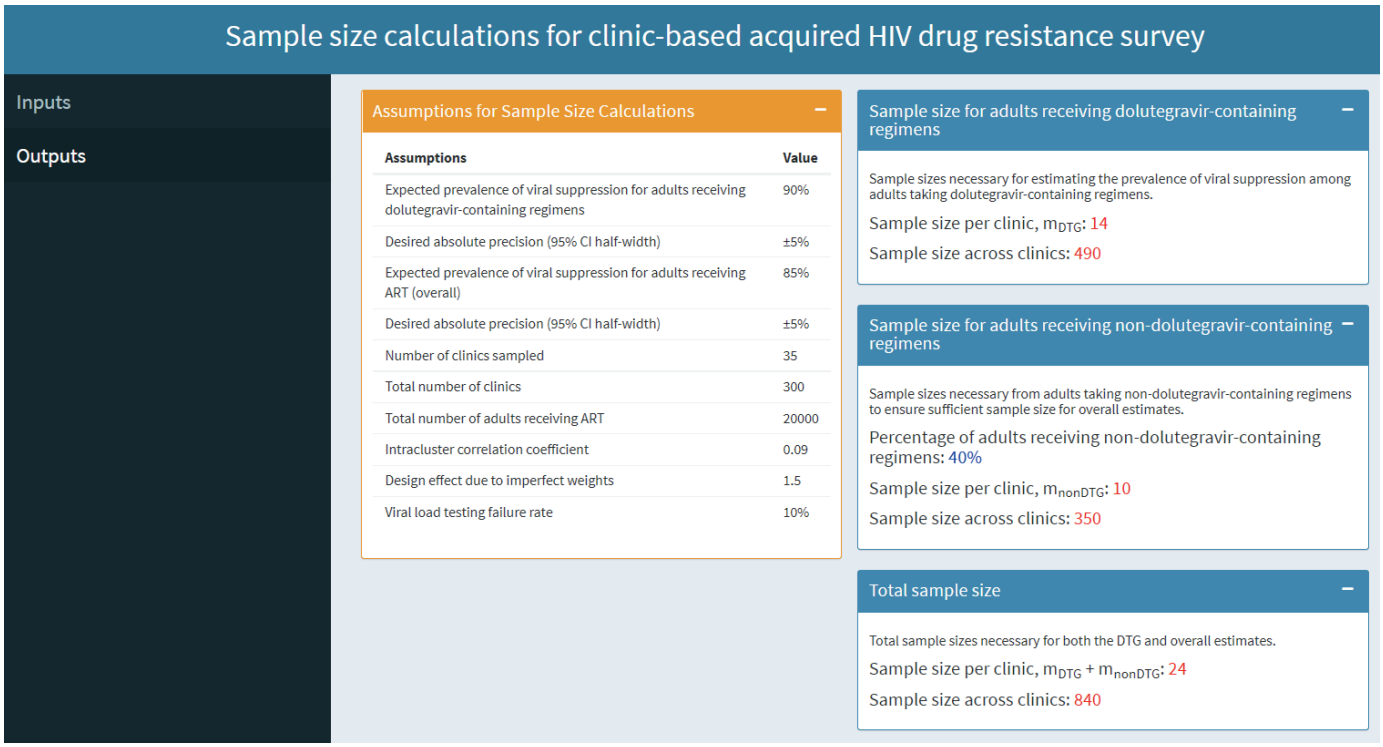
Fig. A3.2. Online sample size calculator: example number of clinics to be sampled



Outputs

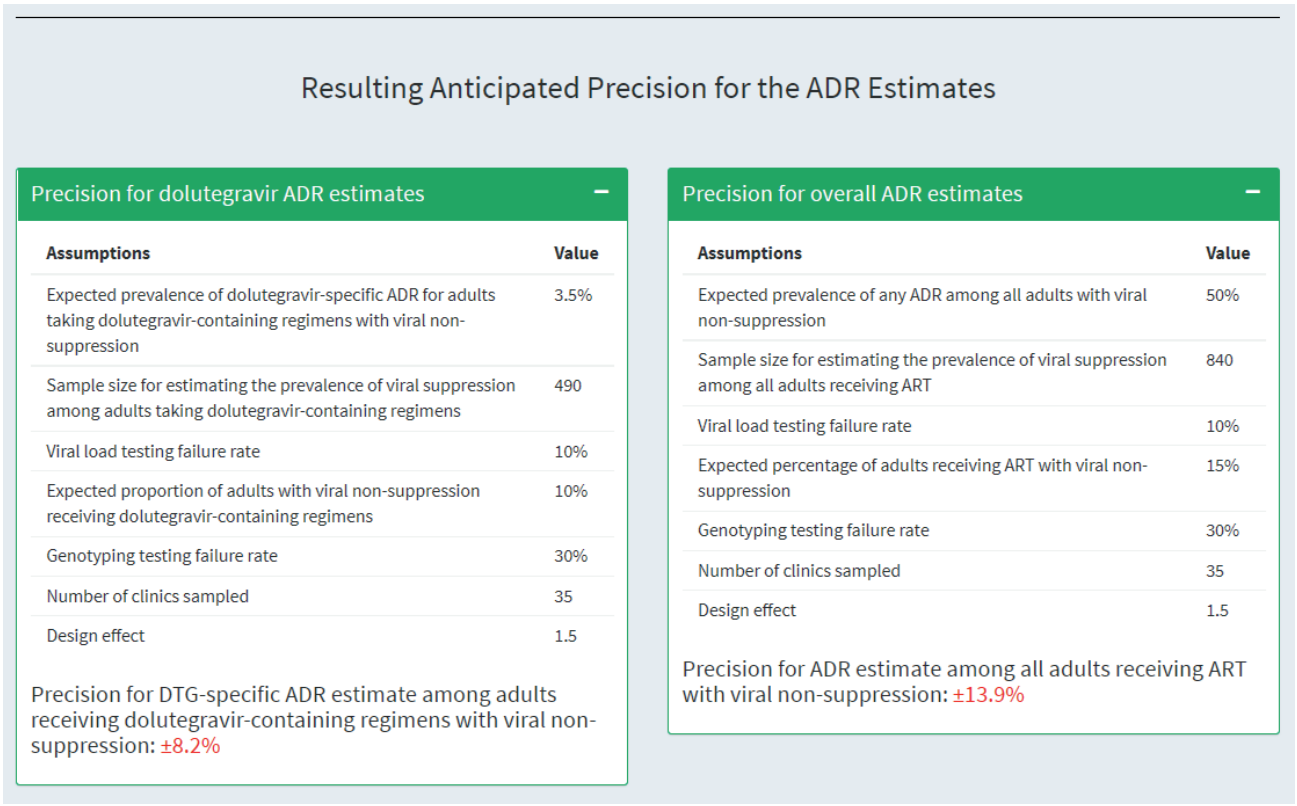
WHO's online sample size calculator automatically calculates all of the details for sample size calculations. The user can then click on the "Outputs" tab on the side panel. In Fig. A3.3 the assumed parameter values for the sample size calculations are displayed in the orange box on the top left. On the top right, there are DTG, non-DTG and total per-clinic and across-clinic sample sizes. For this example, each clinic must sample 14 specimens from adults receiving DTG-containing regimens and 10 specimens from adults receiving non-DTG-containing regimens, for a combined target sample size of 24 specimens per clinic and 840 specimens total, distributed across the 35 selected clinics.

Fig. A3.3. Online sample size calculator: example output of sample size of specimens from adults receiving DTG-containing ART, adults receiving non-DTG-containing ART and the total sample size



The bottom half of the "Outputs" page displays anticipated precision for the prevalence of DTG-specific ADR for individuals receiving DTG-containing regimens with VNS and for the prevalence of any ADR for all people with VNS (see Fig. A3.4). These estimates are calculated using the sample sizes shown in the top half of the page, combined with a number of assumed parameter values. For this example, the precision for DTG-specific acquired HIV drug resistance is 8.2%, and the precision for any acquired HIV drug resistance is 13.9%.

Fig. A3.4. Online sample size calculator: example output of anticipated precision for DTG-specific ADR and any ADR



Resulting Anticipated Precision for the ADR Estimates

Precision for dolutegravir ADR estimates

Assumptions	Value
Expected prevalence of dolutegravir-specific ADR for adults taking dolutegravir-containing regimens with viral non-suppression	3.5%
Sample size for estimating the prevalence of viral suppression among adults taking dolutegravir-containing regimens	490
Viral load testing failure rate	10%
Expected proportion of adults with viral non-suppression receiving dolutegravir-containing regimens	10%
Genotyping testing failure rate	30%
Number of clinics sampled	35
Design effect	1.5

Precision for DTG-specific ADR estimate among adults receiving dolutegravir-containing regimens with viral non-suppression: ±8.2%

Precision for overall ADR estimates

Assumptions	Value
Expected prevalence of any ADR among all adults with viral non-suppression	50%
Sample size for estimating the prevalence of viral suppression among all adults receiving ART	840
Viral load testing failure rate	10%
Expected percentage of adults receiving ART with viral non-suppression	15%
Genotyping testing failure rate	30%
Number of clinics sampled	35
Design effect	1.5

Precision for ADR estimate among all adults receiving ART with viral non-suppression: ±13.9%

ANNEX 4 BUDGET CONSIDERATIONS

This annex provides two example budgets for planning purposes. The first is a generic budget for a clinic-based acquired HIV drug resistance survey among adults receiving ART (Table A4.1) and the second is a generic budget for a clinic-based acquired HIV drug resistance survey among adults and children and adolescents receiving ART (Table A4.2).

Table A4.1. Generic budget for clinic-based acquired HIV drug resistance survey among adults receiving ART (in US dollars)

Example	Total number of ART clinics in the country: 800	Total number of ART clinics sampled: 30	Total number of adults receiving ART: 700 000	Sample size: 900

Protocol development and training

	Number of staff per site	Transport costs	Per diem cost	Number of nights	Total
Training of site staff (1-day training)	2	200.00	150.00	1	21 000.00
Production of protocol and training materials					15 000.00
				Subtotal	**36 000.00**

Survey coordination

	Number of staff	Cost per staff/ month	Number of months	Number of sites	Total
Site coordination	1	300.00	5	30	45 000.00
Nurse incentive	2	50.00	4	30	12 000.00
National coordination	1	1 000.00	12		12 000.00
Data manager	1	800.00	8		6 400.00
				Subtotal	**75 400.00**

Site support visits

	Total
Study coordinator and driver (2 days per visit, US$ 50 per diem, 2 visits)	12 000.00
Fuel (for 6 months)	2 000.00
Air tickets to remote sites (5 flights, US$ 200 each)	1 000.00
Local transportation	1 000.00
Subtotal	**16 000.00**

Laboratory

	Cost per unit	Total
Blood collection	3.00	2 700.00
Dried blood spot preparation and storage	5.00	4 500.00
Viral load testing	60.00	54 000.00
Genotyping for reverse transcriptase, protease and integrase; costs including labour	150.00[a]	33 750.00[b]
Local shipment of specimens (US$ 100 per site for national shipping)		3 000.00
Shipment of specimens to a WHO-designated laboratory (outside the country)		5 000.00
Subtotal		**102 950.00**

Technical support

	Total
Consultant and protocol development, data analysis and report writing and flight (US$ 550 for 20 days and daily per diem USD$ 200 for 7 days); international flight USD$ 3000	15 400.00
Statistical consultant – support statistical analysis (USD$ 550 per day for 12 days)	6 600.00
Subtotal	**22 000.00**

Report production, printing and distribution

	Total
Report production and distribution	10 000.00
Workshop to discuss policy implications and actions required (15 outside participants, 15 local)	10 000.00
Subtotal	**20 000.00**
Total	**272 350.00**

[a] The cost of the HIVDR test should be adapted based on the laboratory quotation (ranging from US$ 50 to 350 per test).
[b] Assuming 25% of individuals enrolled with viral load ≥1000 copies/mL.

Table A4.1. Generic budget for clinic-based acquired HIV drug resistance survey among adults receiving ART (in US dollars)

Example	Total number of ART clinics in the country: 800	Total number of ART clinics sampled: 30	Total number of adults receiving ART: 700 000	Sample size: 900
	Total number of ART clinics in the country for children and adolescents: 600		Total number of children and adolescents receiving ART: 45 000	Children and adolescent sample size: 1200

						Shared survey costs	Survey among adults	Survey among children and adolescents
Protocol development and training								
	Number of staff per site	**Transportation costs**	**Per diem cost**	**Number of nights**	**Total**			
Training of site staff (1-day training)	2	200.00	150.00	1	21 000.00			
Production of protocol and training materials					15 000.00			
				Subtotal	**36 000.00**			
Survey coordination								
	Number of staff	**Cost per staff/month**	**Number of months**	**Number of sites**	**Total**			
Site coordination	1	300.00	5	30	45 000.00			
Nurse incentive	2	50.00	4	30	12 000.00			
National coordination	1	1 000.00	12		12 000.00			
Data manager	1	800.00	8		6 400.00			
				Subtotal	**75 400.00**			
Site support visits								
					Total			
Study coordinator and driver (2 days per visit, US$ 50 per diem, 2 visits)					12 000.00			
Fuel (for 6 months)					2 000.00			
Air tickets to remote sites (5 flights, US$ 200 each)					1 000.00			
Local transportation					1 000.00			
				Subtotal	**16 000.00**			
Laboratory								
				Cost per unit	**Total**	**Total**	**Total**	
Blood collection				3.00		2 700.00	3 600.00	
Dried blood spot preparation and storage				5.00		4 500.00	6 000.00	
Viral load testing				60.00		54 000.00	72 000.00	
Genotyping for reverse transcriptase, protease and integrase; costs including labour				150.00[a]		33 750.00[b]	45 000.00[b]	
Local shipment of specimens (US$ 100 per site for national shipping)					3 000.00			
Shipment of specimens to a WHO-designated laboratory (outside the country)					5 000.00			
				Subtotal	**8 000.00**	**94 950.00**	**126 600.00**	
Technical support								
					Total			
Consultant and protocol development, data analysis and report writing and flight (US$ 550 for 20 days and daily per diem USD$ 200 for 7 days); international flight USD$ 3000					15 400.00			
Statistical consultant – support statistical analysis (USD$ 550 per day for 12 days)					6 600.00			
				Subtotal	**22 000.00**			
Report production, printing and distribution								
					Total			
Report production and distribution					10 000.00			
Workshop to discuss policy implications and actions required (15 outside participants, 15 local)					10 000.00			
				Subtotal	**20 000.00**			
				Total	**398 950.00**			

[a] The cost of the HIVDR test should be adapted based on the laboratory quotation (ranging from US$ 50 to 350 per test).
[b] Assuming 25% of individuals enrolled with viral load ≥1000 copies/mL.

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 978-92-4-003500-3