

HIV DRUG RESISTANCE

CONCEPT NOTE

# SURVEILLANCE OF HIV DRUG RESISTANCE IN ADULTS RECEIVING ART (ACQUIRED HIV DRUG RESISTANCE)

JULY 2014





# **SURVEILLANCE OF HIV DRUG RESISTANCE IN ADULTS RECEIVING ART (ACQUIRED HIV DRUG RESISTANCE)**

JULY 2014

WHO Library Cataloguing-in-Publication Data

Surveillance of HIV drug resistance in adults receiving ART (acquired HIV drug resistance).

1. Anti-HIV agents – therapeutic use. 2. Drug resistance, Viral – drug effects. 3. HIV infections – epidemiology. 4. HIV infections – drug therapy 5. Population surveillance. I. World Health Organization.

ISBN 978 92 4 150707 3

(NLM classification: QV 268.5)

© World Health Organization 2014

All rights reserved. Publications of the World Health Organization are available on the WHO website ([www.who.int](http://www.who.int)) or can be purchased from WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (tel.: +41 22 791 3264; fax: +41 22 791 4857; e-mail: [bookorders@who.int](mailto:bookorders@who.int)).

Requests for permission to reproduce or translate WHO publications –whether for sale or for non-commercial distribution– should be addressed to WHO Press through the WHO website ([www.who.int/about/licensing/copyright\\_form/en/index.html](http://www.who.int/about/licensing/copyright_form/en/index.html)).

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by the World Health Organization to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the World Health Organization be liable for damages arising from its use.

Layout by L'IV Com Sàrl, Villars-sous-Yens, Switzerland.

Printed by the WHO Document Production Services, Geneva, Switzerland.

# TABLE OF CONTENTS

Acknowledgments .....	3
Acronyms .....	4
Background .....	5
1. Introduction .....	6
2. Objectives .....	6
3. Overview of survey approach .....	7
3.1 Survey methodology .....	7
3.2 Survey populations .....	7
4. Survey outcomes .....	8
4.1 Outcome 1a: Prevalence of viral load suppression (VL<1000 copies/mL) among individuals on ART .....	8
4.2 Outcome 1b: Prevalence of viral load suppression (VL<1000 copies/mL) among individuals on first-line ART .....	8
4.3 Outcome 1c: Prevalence of viral load suppression (VL<1000 copies/mL) among individuals on NNRTI-based first-line ART .....	8
4.4 Outcome 2a: Nationally representative measure of retention at 12 months .....	9
4.5 Outcome 2b: Prevalence of viral load suppression (VL<1000 copies/mL) among individuals receiving ART for 12 ( $\pm$ 3) months, adjusted for retention (or worst-case estimate of VL suppression) .....	9
4.6 Outcome 3a: Prevalence of HIVDR among individuals on ART with VL $\geq$ 1000 copies/mL .....	10
4.7 Outcome 3b: Prevalence of HIVDR among individuals receiving first-line ART with VL $\geq$ 1000 copies/mL .....	10
4.8 Outcome 3c: Prevalence of HIVDR among individuals receiving NNRTI-based first-line ART with VL $\geq$ 1000 copies/mL .....	10
4.9 Outcome 4: Prevalence of HIVDR among individuals on ART .....	11
5. How to select clinics .....	12
5.1 Sampling very small or difficult-to-access clinics .....	12
5.2 Regional representation .....	13
5.3 Countries with many ART clinics .....	13
6. Survey among adults receiving ART for 12 ( $\pm$ 3) months .....	14
6.1 Patient eligibility criteria .....	14
6.1.1 Inclusion criteria .....	14
6.1.2 Exclusion criterion .....	14
6.2 Defining the survey sample size .....	14
6.2.1 Assumptions .....	14
6.2.2 Sample size calculations .....	14
6.2.3 Countries with few ART clinics .....	15

7. Survey among adults receiving ART for at least 48 months	16
7.1 Patient eligibility criteria	16
7.1.1 Inclusion criteria	16
7.1.2 Exclusion criterion	16
7.2 Defining the survey sample size	16
7.2.1 Assumptions	16
7.2.2 Sample size calculations	16
7.2.3 Countries with few ART clinics	17
8. Laboratory methods	18
8.1 Specimen collection, handling, processing and tracking	18
8.2 HIVDR genotyping and quality assurance of sequences	18
9. Implementation considerations	19
9.1 Duration of the survey, patient screening and sampling	19
9.2 List of variables to be collected	19
9.2.1 Patient-level information	19
9.2.1.1 Patient-level epidemiological and laboratory variables to be captured for all patients who will have blood drawn for VL assessment and genotyping	19
9.2.1.2 Minimum information to be captured for all eligible patients for at least the first three months from survey initiation	19
9.2.2 Clinic-level information	20
9.2.3 Survey-level information	21
9.3 Patient under-enrolment	21
9.4 Repeating the survey	21
9.5 Tools for country adaptation	21
9.6 Combining acquired HIVDR surveys with surveys of pre-treatment HIVDR	21
10. Data analysis	23
ANNEXES	24
Annex 1.1: Selecting the clinics to survey	24
Annex 1.2: Guaranteeing representation from all regions	26
Annex 1.3: Stratification	26
Annex 1.4: Data analysis plan	27
Annex 1.5: Reporting of HIVDR data	35
Annex 1.6: Calculating a nationally representative estimate of retention	35
Statistical Appendix	38
1. Sample size calculations: Prevalence of viral load suppression	38
2. Calculating the contribution to the design effect due to clustering of the outcome by clinic	39
3. Calculating the contribution to the design effect due to imperfect weighting information	40
4. Other sources of design effect	40
5. Calculating the sample size	40
6. Incorporating the finite population correction	41
7. Sample size calculations when all clinics are sampled	42
8. Sample size calculations for nationally representative estimate of retention	42
9. Estimated confidence interval width for Outcome 2b	43
10. Data analysis	43

# ACKNOWLEDGMENTS

The work was led by Silvia Bertagnolio (World Health Organization), Jhoney Barcarolo (consultant) and Michael Jordan (Tufts University School of Medicine, USA). Statistical methods were developed by Natalie Exner Dean and Marcello Pagano (Harvard University, USA). Natalie Exner Dean wrote the statistical annexes and analysis plan. The concept note received substantial technical inputs and revisions by the HIVResNet Core Group (Emiliano Bissio, Argentina; Avelin Aghokeng, Cameroun; Elliot Raizes, USA; and Andrea De Luca, Italy).

Jos Perriens, coordinator of the Technologies and Commodities Team at the HIV Department, World Health Organization, provided overall supervision. We thank the United States Centers for Disease Control and Prevention, Atlanta, USA and the President's Emergency Plan for AIDS Relief (PEPFAR) HIV Drug Resistance working group for their critical inputs to the document.

We are most grateful to the HIVResNet Steering Group (Chunfu Yang, USA; Chris Archibald, Canada; Achara Teeraratkul, Thailand; Avelin Aghokeng, Cameroun; AM.J. Wensing, the Netherlands; Christopher Duncombe, USA; Kerry Dierberg, USA; Do Thi Nhan, Vietnam; Emiliano Bissio, Argentina;

Elliot Raizes, USA; Jonathan Shapiro, Israel; Natalia Nizova, Ukraine; Robert W. Shafer, USA; Wilford Kirungi, Uganda; Zhang Fujie, China; Tobias Rinke de Wit, the Netherlands; and Sergio Carmona, South Africa) for their strategic advice. We are also grateful to the contributions provided by WHO staff from regional and country offices.

Additionally, development of this guidance document benefitted greatly from inputs received from country ART programme managers and country representatives attending WHO regional workshops conducted in 2013 in Beijing, China; Brasilia, Brazil; Montpellier, France; Addis Ababa, Ethiopia; and Cape Town, South Africa. WHO is grateful to Amandine Cournil, Montpellier, France, for her review of the statistical annexes. Development and publication of this guidance was supported by the Bill and Melinda Gates Foundation (#38180); we are sincerely grateful for their support.

# ACRONYMS

<b>ADR</b>	acquired HIV drug resistance
<b>AIDS</b>	acquired immunodeficiency syndrome
<b>ART</b>	antiretroviral therapy
<b>ARV</b>	antiretroviral drug
<b>ATV/r</b>	atazanavir/ritonavir
<b>AZT</b>	zidovudine (also ZDV)
<b>DEFF</b>	design effect
<b>DBS</b>	dried blood spot
<b>DRV/r</b>	darunavir/ritonavir
<b>EFV</b>	efavirenz
<b>HIV</b>	human immunodeficiency virus
<b>ICC</b>	intracluster correlation coefficient
<b>HIVDR</b>	HIV drug resistance
<b>INI</b>	integrase inhibitor
<b>LPV/r</b>	lopinavir/ritonavir
<b>mL</b>	milliliter
<b>NNRTI</b>	non-nucleoside reverse transcriptase inhibitor
<b>NRTI</b>	nucleoside reverse transcriptase inhibitor
<b>N(t)RTI</b>	nucleotide reverse transcriptase inhibitor
<b>NVP</b>	nevirapine
<b>PDR</b>	pretreatment HIV drug resistance
<b>PEPFAR</b>	President's Emergency Plan for AIDS Relief
<b>PI</b>	protease inhibitor
<b>PPS</b>	probability proportional to size
<b>PPPS</b>	probability proportional to proxy size
<b>PR</b>	protease region
<b>RT</b>	reverse transcriptase region
<b>SI</b>	sampling interval
<b>SID</b>	survey identification number
<b>VL</b>	viral load
<b>UNAIDS</b>	Joint United Nations Programme on HIV/AIDS
<b>UNGASS</b>	United Nations General Assembly Special Session
<b>WHO</b>	World Health Organization



# BACKGROUND

HIV Drug Resistance emerges when HIV replicates in the presence of antiretroviral drugs. If HIV drug resistance becomes widespread, the drugs currently used to treat HIV infection may become ineffective. To date, levels of HIV Drug Resistance in countries scaling up ART remain manageable. However, resistance is slowly increasing: in East Africa, resistance rates of 10% to non-nucleoside drugs (such as nevirapine and efavirenz) have been recently described.

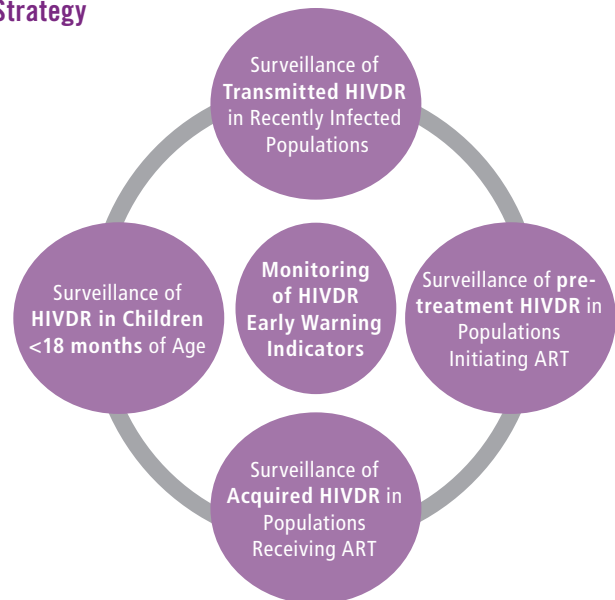
To maximize the long-term effectiveness of first-line ART regimens, and ensure the sustainability of ART programmes, it is essential to minimize the further spread of HIV drug resistance. Even in settings with optimal ART programme management, some degree of HIVDR is expected to emerge in populations on ART and some HIVDR is expected to be transmitted to previously uninfected individuals. Therefore, WHO recommends that HIV treatment scale-up should always be accompanied by a robust assessment of drug resistance emergence and transmission. WHO's HIVDR Monitoring and Surveillance Strategy is composed of five key elements:

- i. Monitoring of Early Warning Indicators of HIV drug resistance
- ii. Surveillance of HIVDR in recently-infected adult populations (transmitted HIVDR)
- iii. Surveillance of pre-treatment HIVDR in adult populations initiating ART (pre-treatment HIVDR)
- iv. Surveillance of acquired HIVDR in populations of adults and children receiving ART (acquired HIVDR)
- v. Surveillance of HIV drug resistance in treatment-naive children less than 18 months of age

WHO's HIVDR Monitoring and Surveillance Strategy is a critical component of the public health approach to ART delivery. By obtaining population-level data on HIVDR in different populations, its various elements can inform programme-level decision making regarding, for example, optimal first and second lines, for both children and adults. Management of treatment failure (i.e. what to do when a patient fails a particular regimen) is addressed in the guidance note on ART.

This document describes methods to assess HIVDR in adult populations on ART (surveillance of acquired HIVDR).

**Figure 1. HIV Drug Resistance Monitoring and Surveillance Strategy**



# 1. INTRODUCTION

Impressive gains have been made over the last 10 years in expanding access to antiretroviral therapy (ART) in low- and middle-income countries. From fewer than 300,000 in December 2003, there are now over 9.7 million people receiving ART in resource-limited settings. While historically the focus of the global HIV response has been on rapid scale-up, millions of people have now been on ART for a considerable period. Therefore, it has become increasingly important to assess, in a standardized and nationally representative manner, the extent to which, at different points over time, those receiving ART achieve viral load (VL) suppression and the extent to which HIV drug resistance (HIVDR) is emerging among individuals failing ART.

Results from surveys to assess acquired drug resistance (ADR) provide critical information to assess the performance

of programmes in maximizing viral suppression, inform the optimal selection and management of second-line therapies, and provide insight on the extent to which patients are switching therapies unnecessarily. Armed with results from ADR surveys, programmes can identify gaps in service delivery and implement appropriate policy responses to improve individual and population outcomes.

This concept note describes methods to assess nationally representative levels of viral load suppression and drug resistance in adults receiving ART through the implementation of a cross-sectional survey. Surveillance of acquired drug resistance in children will be conducted separately and is addressed in a distinct concept note.

# 2. OBJECTIVES

The main purpose of this survey is to calculate nationally representative prevalence estimates (with associated confidence intervals) of (1) VL suppression and (2) of HIVDR in populations receiving ART for 12 ( $\pm 3$ ) months and for  $\geq 48$  months.

## 3. OVERVIEW OF SURVEY APPROACH

### 3.1 Survey methodology

This survey uses a method known as a two-stage cluster design. In the first stage, a minimum of 17 clinics are sampled from a list of all clinics dispensing ART in the country. In the second stage, a sample of eligible patients is recruited from each of the selected clinics. Patients included in the survey will have blood specimens collected for VL testing. Specimens with a VL of  $\geq 1000$  copies/mL will be genotyped to determine HIVDR status.

For the purpose of this concept note the words “clinic” and “site” are used interchangeably.

### 3.2 Survey populations

The survey can be performed in populations receiving ART at two different timepoints:

1. Early timepoint – targeting adults who have been on ART for 12 ( $\pm 3$ ) months
2. Late timepoint – targeting adults who have been on ART for at least 48 months

Considering that patients do not visit clinics necessarily at 12 months specifically, a margin of  $\pm 3$  months has been added to

the first timepoint to improve its feasibility. Moreover, similar outcomes with respect to VL suppression, HIVDR and retention are likely to be observed among patients in this relatively short time bracket (9–15 months)<sup>1,2</sup>.

Understanding VL and HIVDR levels among populations at an early timepoint is essential to learn whether there are programme gaps to be addressed in order to ensure optimal treatment outcomes. Assessment of VL in adults at a later timepoint investigates whether population-level VL suppression has been maintained for periods greater than one year and provides an opportunity to characterize HIVDR in this population. In particular, adults with detected HIVDR who have been on ART for at least 48 months may have been on a failing regimen for longer, and therefore may have different mutations or mutation patterns (specifically an accumulation of thymidine analogue mutations), than adult populations found to be failing ART 12 months after initiation. Moreover, evidence suggests that, of those patients who switch to second-line therapies, most have been on first-line ART for at least 48 months, so that assessing HIVDR at this point can help optimize subsequent regimen choice.<sup>3</sup>

Each timepoint requires a different sample with distinct sample sizes (see Sections 6 and 7). Countries can choose to perform the surveys separately or jointly for the two timepoints.

1 Global update on HIV treatment 2013. Geneva, World Health Organization, 2013.

2 Barth et al. (2011) Long-term outcome of an HIV-treatment programme in rural Africa: viral suppression despite early mortality. *AIDS Research and Treatment*, Volume 2011, Article ID 434375.

3 Paton N. et al. (2013) Presentation on Africa Research Network for Evaluation of Second-line Therapy (EARNEST). International AIDS Conference 2013, Kuala Lumpur ([pag.ias2013.org/PAGMaterial/PPT/611\\_578/final3.pptx](http://pag.ias2013.org/PAGMaterial/PPT/611_578/final3.pptx), accessed 6 September 2013).

## 4. SURVEY OUTCOMES

The population-level outcomes of this survey are described in Table 1 below. They will be analysed separately for each of the two timepoints.

**Table 1: Summary of survey outcomes**

	Outcomes	Survey early timepoint 12 ( $\pm$ 3) months	Survey late timepoint $\geq$ 48 months
1a	Prevalence of VL suppression (VL<1000 copies/mL) among individuals on ART	X	X
1b	Prevalence of VL suppression (VL<1000 copies/mL) among individuals on first-line ART	X	X
1c	Prevalence of VL suppression (VL<1000 copies/mL) among individuals on NNRTI-based first-line ART	X	X
2a	Nationally representative measure of retention at 12 months	X	
2b	Prevalence of VL suppression among individuals on ART, adjusted for retention	X	
3a	Prevalence of HIVDR among individuals on ART with VL $\geq$ 1000 copies/mL	X	X
3b	Prevalence of HIVDR among individuals on first-line ART with VL $\geq$ 1000 copies/mL	X	X
3c	Prevalence of HIVDR among individuals on NNRTI-based first-line ART with VL $\geq$ 1000 copies/mL	X	X
4	Prevalence of HIVDR among individuals on ART	X	X

NNRTI= non-nucleoside reverse transcriptase inhibitor

### 4.1 Outcome 1a: Prevalence of viral load suppression (VL<1000 copies/mL) among individuals on ART

Outcome 1a estimates the proportion of individuals sampled achieving viral load suppression (VL<1000 copies/mL) among all patients sampled with VL testing successful and results classifiable. It provides a prevalence estimate with an associated 95% confidence interval.

### 4.2 Outcome 1b: Prevalence of viral load suppression (VL<1000 copies/mL) among individuals on first-line ART<sup>1</sup>

Outcome 1b measures the proportion of patients sampled receiving *first-line* ART achieving viral load suppression (VL<1000 copies/mL) among all individuals sampled receiving first-line ART with VL testing successful and results classifiable.

It is important to assess VL suppression while on first-line ART as an important programme objective is to maximize the duration of first-line ART.

In countries where all patients sampled are receiving first-line ART, Outcomes 1a and 1b will be the same. First line is defined as per national treatment guidelines and may include non-nucleoside reverse transcriptase (NNRTI)-based, protease inhibitor (PI)-based or other regimens.

### 4.3 Outcome 1c: Prevalence of viral load suppression (VL<1000 copies/mL) among individuals on NNRTI-based first-line ART

Outcome 1c measures the proportion of patients sampled receiving *NNRTI-based first-line* ART achieving viral load suppression (VL<1000 copies/mL) among all individuals sampled receiving *NNRTI-based* first-line ART with VL testing successful and results classifiable.

In some countries with more mature ART programmes, such as those in Latin America and the Caribbean, where ART has been available for over a decade, an important proportion of patients on ART now receive non-NNRTI-based combinations,

<sup>1</sup> The first-line regimen may be the first ART regimen which the patient was prescribed, or it may be an alternative first-line regimen that was started as a substitution. A patient is still classified as being on first line if his/her treatment has been changed from one first-line ART regimen to another first-line ART regimen (intra-class substitution), for example due to adverse events or toxicity.

such as protease inhibitors. In such cases, this outcome is critical to inform programme decision making. Compared to Outcome 1b, the sample size for Outcome 1c should be further adjusted to account for the proportion of individuals receiving non-NNRTI-based regimens.

In countries where all patients starting ART are prescribed an NNRTI-based regimen, this outcome will be identical to Outcome 1b.

#### 4.4 Outcome 2a: Nationally representative measure of retention at 12 months

Unlike a cohort study, through which a group of patients is assessed continuously over time, a cross-sectional survey, by its very nature, excludes patients who are no longer receiving ART and therefore cannot be observed because they have died, been lost to follow-up or have stopped treatment. This survivor bias can significantly impact the interpretation of Outcome 1a. In order to adjust for such a possible bias, nationally representative data on retention at 12 months are needed.

Outcome 2a is a nationally representative measure of retention at 12 months, and it will be used to adjust the observed prevalence of viral load suppression among individuals on ART for the proportion of individuals who are no longer attending clinics and thus cannot be sampled.

It is critical to ensure that all retention data used to calculate Outcome 2a are of high quality and nationally representative to support adequate interpretation for policy-making. Thus, two methods may be used to calculate and report this outcome:

1. *Census*: international guidance recommends that countries assess 12-month retention by reviewing outcomes of *all* adults who initiated ART in a particular reporting period across all clinics<sup>1,2</sup>. In countries where such a census is performed, the resulting national retention rate can be used to report Outcome 2a.
2. *Survey*: in countries where not all patients or clinics can be included in the retention assessment, Annex 1.6 provides instructions on how to calculate a nationally representative estimate of retention by extracting information from the same clinics selected for the survey.

Outcome 2a should not be reported if the national retention assessment is not based on a census or on the methods described in Annex 1.6.

1 Global AIDS response progress reporting 2012: Guidelines for construction of core indicators for monitoring the 2011 political declaration on HIV/AIDS. Geneva, UNAIDS, 2011.

2 The President's Emergency Plan for AIDS Relief. Next generation indicators reference guide. Version 1.1, 2009.

#### 4.5 Outcome 2b: Prevalence of viral load suppression (VL<1000 copies/mL) among individuals receiving ART for 12 (±3) months, adjusted for retention (or worst-case estimate of VL suppression)

As previously discussed, without accounting for differences in retention within and across countries, it is challenging to meaningfully interpret Outcomes 1a, 1b and 1c. Therefore, in order to assess changes in the national estimate of viral load suppression over time, to compare these estimates against a global standard, or to compare estimates across countries, Outcome 2b is a measure of viral load suppression that combines the observed cross-sectional data with data on patient retention on ART.

Data on VL or HIVDR outcomes among non-retained individuals (that is, those who are lost to follow-up, stop therapy or die after ART initiation) are scarce. There is evidence that a proportion of individuals considered to be lost to follow-up have in fact silently transferred to another clinic and are therefore still on treatment elsewhere. One systematic review has found that 33–48% of the people lost to follow-up after ART initiation had in fact died, and a further 12–54% of those lost to follow-up had silently self-transferred to another facility and were accessing care elsewhere<sup>3</sup>. In South Africa, a study showed that 13% of the people initiating ART had transferred-out within 2.5 years<sup>4</sup>. Given the difficulties associated with determining the precise outcomes of those individuals who are no longer in care, the VL suppression rate in this group is conservatively assumed to be 0%. In so doing, Outcome 2b represents a worst-case estimate of VL suppression.

In countries estimating national retention at 12 months following the methods described in Annex 1.6, the adjusted virological suppression estimate is performed by (1) multiplying clinic-specific data on retention by clinic-specific data on unadjusted virological suppression and (2) aggregating results across clinics. Additional information can be found in Annex 1.6 and in the Statistical Appendix. The width of the confidence interval of the adjusted VL suppression rate will reflect the desired confidence intervals for the unadjusted prevalence of VL suppression and for the retention measure.

Outcome 2b should not be reported if the retention measure used is not based on a census or following the methods described in Annex 1.6.

3 Brinkhof MW, Pujades-Rodriguez M, Egger M (2009). Mortality of patients lost to follow-up in antiretroviral treatment programmes in resource-limited settings: systematic review and meta-analysis. *PLoS One*, 4:e5790.

4 Nglazi MD et al. (2013). Increasing transfers-out from an antiretroviral treatment service in South Africa: patient characteristics and rates of virological non-suppression. *PLoS One*, 8:e57907.

## Box 1: Interpreting Outcomes 2a and 2b – an example

Suppose Outcome 1a (that is, the unadjusted prevalence of viral load suppression among retained patients) in Country A is 75%, and 80% in Country B. A direct comparison of viral load suppression rates among retained patients would suggest that Country B performs better relative to Country A. However, programme data reveal that, in Country A, of the people who initiated therapy approximately 12 months ago, 90% have been retained, while in Country B, retention is only 60%.

In this case, the adjusted prevalence of viral load suppression for Country A would be approximately 67% (see Annex 1.4 for detailed instructions on data analysis). However, for Country B, this would be approximately 48%. This comparison highlights that even though Country B has a higher Outcome 1a, due to poor retention, the national prevalence of VL suppression of individuals on ART for 12 months is likely higher in Country A than in Country B. While Outcome 2b provides an aggregate figure that summarizes programme performance in achieving VL suppression and allows for in-country comparisons to be performed over time, it should be interpreted in the context of its components.

Due to the open-ended nature of the late timepoint, no retention estimation is performed for individuals receiving ART for  $\geq 48$  months, and Outcome 2b will only be calculated for the early timepoint.

### 4.6 Outcome 3a: Prevalence of HIVDR among individuals on ART with VL $\geq 1000$ copies/mL

Outcome 3a estimates the proportion of individuals sampled with VL  $\geq 1000$  copies/mL and detected drug resistance<sup>1</sup> among all individuals sampled with VL  $\geq 1000$  copies/mL and successfully genotyped.

The survey is not designed to achieve a particular confidence interval width for this outcome due to the very large sample size that would be required given that only patients with VL failure are eligible for inclusion in Outcome 3a.

### 4.7 Outcome 3b: Prevalence of HIVDR among individuals receiving first-line ART with VL $\geq 1000$ copies/mL

Outcome 3b estimates the proportion of individuals sampled with VL  $\geq 1000$  copies/mL with detected drug resistance<sup>1</sup> receiving first-line ART among all individuals sampled with VL  $\geq 1000$  copies/mL receiving first-line ART and successfully genotyped.

This distinction is important as the prevalence and patterns of HIVDR among patients failing first or second lines are likely to be different, and accurate measures are needed

to inform optimal regimen selection. Importantly, as it is a sub-analysis of Outcome 3a, no additional data collection is needed to calculate Outcome 3b. First line is defined as per national treatment guidelines and may include NNRTI-based, PI-based or other regimens. In countries where all patients sampled receive a first-line combination, Outcomes 3a and 3b will be the same.

The survey is not designed to achieve a particular confidence interval width for Outcome 3b for the same reasons discussed in the context of Outcome 3a. However, at the analysis stage, data can be aggregated at the regional or global levels to obtain a sufficiently large sample size and produce an estimate with an acceptable confidence interval.

### 4.8 Outcome 3c: Prevalence of HIVDR among individuals receiving NNRTI-based first-line ART with VL $\geq 1000$ copies/mL

Outcome 3c estimates the proportion of individuals sampled with VL  $\geq 1000$  copies/mL with detected drug resistance<sup>2</sup> receiving NNRTI-based first-line ART among all individuals sampled with VL  $\geq 1000$  copies/mL receiving NNRTI-based first-line ART and successfully genotyped.

Similarly to Outcome 1c, in countries where a proportion of people initiate first-line ART regimens that are not NNRTI-based, this outcome should be calculated separately and the sample size must be further adjusted to account for the proportion of individuals receiving non-NNRTI-based regimens.

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

<sup>2</sup> NNRTI resistance is defined with respect to NVP, EFV or both. Sequences classified as low-level, intermediate or high-level resistance according to the Stanford HIVdb are aggregated as "HIV drug resistance".



In countries where all patients starting ART are prescribed an NNRTI-based regimen, this outcome will be identical to Outcome 3b.

## 4.9 Outcome 4: Prevalence of HIVDR among individuals on ART

Outcome 4 estimates the proportion of individuals sampled with VL $\geq$ 1,000 copies/mL and detected HIVDR<sup>1</sup> among all individuals sampled with VL testing successful and results classifiable. Contrary to Outcome 1a, the confidence interval is not targeted *a priori*.

All outcomes should be calculated taking into account actual patient accrual rates, potential patient under-enrolment at any clinic, the rate of viral amplification failure by clinic, and genotyping failure. The analysis will account for these elements through adjustments of the survey weights (an example of a data analysis plan is provided in Annex 1.4 and additional technical background is available in the Statistical Appendix).

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

## 5. HOW TO SELECT CLINICS

As discussed in Section 3, this survey approach relies on a method known as a two-stage cluster design. The first stage is the selection of ART clinics where the survey will be conducted. To achieve a nationally representative prevalence estimate, random sampling must be used to select clinics.

The first step in the selection of clinics is to create a *sampling frame*, also called a sampling table, which is a list of the clinics providing ART in the country (as discussed below), alongside the respective number of people on ART in each clinic at the end of the year prior to the survey. In some instances patients are formally linked to a clinic and counted as if attending that clinic, but actually receive care at a satellite health post. In such circumstances, it is important to list such satellite health posts separately in the sampling frame as well, with the actual number of patients attending each of them.

The composition of the clinic list from which survey sites will be selected depends on the populations being surveyed:

1. Countries conducting the survey only among individuals receiving ART for 12 ( $\pm 3$ ) months should list *all* clinics providing ART in the country.
2. Countries conducting the survey only among individuals receiving ART for at least 48 months should *restrict the systematic sampling table to those clinics that have been in operation for at least 48 months*. This will maximize the probability of obtaining the required sample size per clinic.
3. In countries surveying both timepoints concomitantly, the sampling frame will be stratified in two strata by clinic age (for example, all clinics in operation for less than 48 months; and all clinics in operation for at least 48 months). An Excel-based calculator has been developed to assist countries to optimize the distribution of clinics between these two strata and minimize the number of sites that must be sampled in order to achieve the desired confidence intervals for the two timepoints (see Section 9.5).

Operationally, the process for producing the systematic sampling table for clinic selection is described in Annex 1.1. The recommended method to sample clinics is called *probability proportional to proxy size (PPPS) sampling*. In PPPS sampling, clinics are sampled proportional to the total number of patients on ART in each clinic. Clinics with more patients on ART will be more likely to be sampled than smaller clinics. Relevant technical background on PPPS can be found in the Statistical Appendix.

Survey characteristics, including patient eligibility criteria and sample size requirements, differ according to the timepoint of interest (that is, 12 ( $\pm 3$ ) months or  $\geq 48$  months).

### 5.1 Sampling very small or difficult-to-access clinics

Some countries may have a number of clinics with extremely small populations of patients on ART or clinics that may be difficult to access for a variety of reasons, including political instability or geographical remoteness. Although not advisable, some countries may consider excluding some of these clinics from the systematic sampling table due to logistic and under-enrolment issues.

In general, if these clinics represent less than 10% of the population on ART in the country, countries may choose to exclude these clinics from the systematic sampling table. This threshold seeks to limit the potential bias that such exclusion may introduce in the final results. In this case, the 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. On the other hand, if more than 10% of the population of interest attend these clinics, it is not advisable to exclude these clinics from the pool of clinics that can be sampled for the survey.

For example, suppose a country excluded difficult-to-access clinics which represented 10% of the patient population. In this country, the prevalence of viral load suppression among the remaining clinics in the sampling table was 90%. If the prevalence of viral load suppression among the excluded, difficult-to-access clinics was also 90%, then the true prevalence of viral load suppression in the entire population was 90%; thus, excluding these clinics did not introduce bias into the national prevalence estimate. However, if the prevalence of viral load suppression among the excluded, difficult-to-access clinics was much lower, for example, 70%, then the true prevalence of viral load suppression in the entire population was 88%. The survey among the clinics in the sampling table would overestimate the prevalence of viral load suppression (90% versus 88%), though the magnitude of this bias is minor when the percentage of the patient population excluded is low ( $< 10\%$ ) and the difference in prevalence between the included and excluded populations is small.

In general, if the excluded patients have a different prevalence of VL suppression than the observed patients, the national prevalence estimate will be biased.



## 5.2 Regional representation

Countries wishing to develop region-specific estimates of viral load suppression should implement a separate survey in each area of interest. However, if the goal is not to make region-specific inferences, but to balance administrative load and achieve some regional representation, this can be achieved using a technique called implicit stratification by region. Operationally, this can be accomplished by listing clinics by administrative or geographical area prior to their selection (see Annex 1.1). Clinics will be sampled proportionally to the size of the region. It is important to note that this method does not allow for the development of region-specific estimates nor does it always guarantee the geographical representation of all regions. If a country wishes to guarantee the representation of at least one clinic from each area of interest, this can be done using the method described in Annex 1.2.

If knowledge of the prevalence of VL suppression by other characteristics (for example, rural versus urban) is relevant for national decision-making, countries may consider *formal stratification* at the design phase (see Annex 1.3).

## 5.3 Countries with many ART clinics

Countries with a large number of clinics (for example, more than 1000) are not required to sample more than 40 clinics to achieve a nationally representative prevalence estimate assuming that the level of heterogeneity across clinics is consistent with the available global data. Nonetheless, larger countries may have more heterogeneity across clinics than smaller countries. As a result, it is recommended that larger countries sample no fewer than 30 clinics, and they may prefer to sample a greater number for a more precise conclusion if resources are available. Alternatively, if larger countries would like to make region-specific statements with a particular precision, then they should consider conducting a separate survey in each region. These region-specific surveys then typically lead to more precise national numbers when combined across regions.

## 6. SURVEY AMONG ADULTS RECEIVING ART FOR 12 ( $\pm 3$ ) MONTHS

### 6.1 Patient eligibility criteria

#### 6.1.1 Inclusion criteria

- HIV+ adults<sup>1</sup> who provide informed consent, and
- Adults who have been on antiretroviral therapy for 12 ( $\pm 3$ ) months and are still on ART at the time of enrolment<sup>2</sup>, regardless of site of therapy initiation.

#### 6.1.2 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 coinfection are excluded.

### 6.2 Defining the survey sample size

#### 6.2.1 Assumptions

In addition to the clinic sampling method, a number of key model assumptions affect the required survey sample size for Outcomes 1a, 1b and 1c:

1. The expected prevalence of VL suppression among individuals sampled, which is assumed to be 85%.
2. The expected laboratory (VL and genotyping) failure rate, which is assumed to be 15%. Countries should clearly report the number of specimens that fail to yield classifiable VL, those that were not successfully genotyped, and those specimens that were lost, if any.
3. The expected proportion of all individuals sampled still receiving first-line ART. Available evidence suggests that in most countries generally only a small proportion of patients will not be on first-line ART. For the purposes of this exercise, it is assumed that 95% of sampled individuals will still be on a first-line regimen. However, in certain settings, particularly in Latin America and the Caribbean, this proportion may be considerably lower. In order to perform this adjustment, the estimated sample size must be increased to account for the proportion of patients sampled not receiving first-line ART. As countries develop their own national protocols, it is important to adjust sample sizes according to national data on the proportion of individuals on first-line ART.

<sup>1</sup> Adults are generally defined as being 18 years old and above, though the minimum age may be country specific.

<sup>2</sup> This does not require patients to have been on ART continuously for the 12-month period. Therefore, for assessing duration on ART, treatment interruptions should be ignored.

4. The expected proportion of individuals sampled on first-line ART who are receiving NNRTI-based regimens is assumed to be 100%. This assumption reflects the most common situation in countries using the public health approach. However, in some countries, a proportion of patients initiating ART are prescribed non-NNRTI-based combinations (for example, PI-based regimens). In such cases, the sample size must be further inflated to account for the proportion of individuals receiving non-NNRTI-based regimens. For example, if in Country X only 75% of individuals on ART receive an NNRTI-based combination, the relevant sample size should be further inflated by dividing it by 0.75.
5. The desired precision of the estimate: a confidence interval of half-width of  $\pm 5\%$  is suggested as an appropriate compromise between feasibility and precision.

These assumptions are summarized in Table 2.

**Table 2: Key model assumptions to calculate the sample size for Outcomes 1a, 1b and 1c among populations receiving ART for 12 ( $\pm 3$ ) months**

Variable	Assumed value
Expected prevalence of viral load suppression	85%
Expected laboratory (VL and genotyping) failure rate	15%
Expected proportion of individuals sampled still receiving first-line ART	95%
Expected proportion of individuals sampled on first-line ART receiving NNRTI-based regimens	100%
Desired confidence interval half-width	5%

The last three assumptions may be changed as countries adapt this concept note into national protocols.

#### 6.2.2 Sample size calculations

In addition to the assumptions discussed above, the total sample size required is also affected by the number of clinics to be sampled. In general, as more clinics are sampled, better representation of the prevalence of viral load suppression across clinics is achieved. However, for logistic, political or financial reasons, countries may wish to limit the number of

clinics sampled. In this case, it is recommended to sample a minimum of 17–40 clinics to obtain a nationally representative estimate.

Due to evidence that VL suppression outcomes cluster by clinic, it is better to sample more clinics rather than sampling additional patients within a clinic. As more clinics are sampled, better representation of the prevalence of VL suppression across clinics is achieved. We do not, however, observe a continuous drop in the sample size because the calculations stabilize after 30 clinics. The survey design

**Table 3: Standard sample size calculations for surveillance of viral load suppression among individuals receiving ART for 12 ( $\pm 3$ ) months**

Number of clinics to be sampled	Number of samples per clinic	Total sample size
17	28	476
18	26	468
19	24	456
20	23	460
21	21	441
22	20	440
23	19	437
24	18	432
25	18	450
26	17	442
27	16	432
28	15	420
29	15	435
30	14	420
31	14	434
32	13	416
33	13	429
34	12	408
35	12	420
36	12	432
37	11	407
38	11	418
39	11	429
40	10	400

Notes: assumed prevalence of VL suppression among individuals sampled = 85%, confidence interval half-width for Outcome 1a =  $\pm 5\%$ , laboratory failure rate of 15%, proportion of individuals sampled receiving first-line ART = 95%; proportion of individuals on first-line ART receiving NNRTI-based regimens = 100%.

effect is most significant when fewer clinics are sampled (see Statistical Appendix)<sup>1</sup>.

The standard sample sizes presented in Table 3 were calculated assuming an extremely large number of sites and of patients on ART. However, as countries develop their national protocols, sample sizes can be tailored to their local circumstances by using national data on (i) the number of people who initiated ART over a 12 month period approximately 12 months before the survey initiation date (to approximate the number of eligible patients observable during the survey period. For example, if the survey starts in January 2014, this would be the number of people who initiated ART between 01 January 2012 and 31 December 2012<sup>2</sup>) and (ii) the total number of ART sites in the sampling table. The application of such a finite population correction factor will result in a decrease in the estimated total sample size required to reach the same confidence interval<sup>3</sup>.

The figures presented in Table 3 provide an estimate of standard sample sizes (based on the assumptions discussed in Section 6.2.1) that can be useful for budgetary and planning purposes.

### 6.2.3 Countries with few ART clinics

It is recommended that countries with few ART clinics sample all of them. For countries sampling all ART clinics, the standard required total sample size to estimate Outcome 1a (VL suppression) is 364. The number of sampled patients at each clinic should be proportional to the size of the clinic (that is, the number of eligible patients at that clinic). For example, if 25% of patients initiate treatment at one clinic in the country, 25% of the actual sample size should be collected from this clinic. This standard total sample size was calculated assuming an infinitely large population size. Countries can tailor their sample size calculations to their local circumstances by using national data on the number of people who initiated ART 12 months prior to survey initiation (for example, if the survey starts in January 2014, this would be the number of people who initiated ART between 01 January 2012 and 31 December 2012<sup>4</sup>). This adaptation process will result in a decrease in the estimated total sample size required to reach the same confidence interval.

<sup>1</sup> The sample sizes are also impacted by the fact that the per-clinic sample size must be a whole number, and this is then multiplied by the number of clinics to yield the total sample size. Variability in how the per-clinic sample sizes are rounded can result in sample sizes that stay constant or slightly increase with additional clinics.

<sup>2</sup> In the absence of precise figures, this number could be approximated by the difference between the number of people on ART at the end of 2012 and the number of people on ART at the end of 2011.

<sup>3</sup> These adjustments can be performed using the Excel-based calculator discussed in section 9.5.

<sup>4</sup> In the absence of precise figures, this number could be approximated by the difference between the number of people on ART at the end of 2012 and the number of people on ART at the end of 2011.

# 7. SURVEY AMONG ADULTS RECEIVING ART FOR AT LEAST 48 MONTHS

## 7.1 Patient eligibility criteria

### 7.1.1 Inclusion criteria

- HIV+ adults<sup>1</sup> who provide informed consent, and
- Adults who have been on antiretroviral therapy for at least 48 months at the time of clinic visit, regardless of site of therapy initiation, and are still on ART at the time of survey enrolment<sup>2</sup>.

### 7.1.2 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 coinfection are excluded.

## 7.2 Defining the survey sample size

### 7.2.1 Assumptions

The same five assumptions discussed in Section 6.2.1 affect the required survey sample size for estimating viral load suppression among populations receiving ART for at least 48 months. Due to the open-ended nature of the late timepoint, no retention estimate is performed, therefore Outcome 2b is not technically applicable.

With respect to the expected prevalence of viral load suppression, it is assumed that a total of 70% of sampled individuals receiving therapy for at least 48 months will be virologically suppressed. The assumed overall laboratory failure rate is 15%, while the estimated proportion of patients still on first-line ART and the proportion of people on first-line ART who are prescribed NNRTI-based regimens are maintained at 95% and 100%, respectively. Considering these assumptions, a confidence interval of half-width  $\pm 6\%$  is suggested to enhance survey feasibility. The last three assumptions may be changed as countries adapt this concept into national protocols.

### 7.2.2 Sample size calculations

Table 4 provides sample size calculations. As discussed in Section 6.2.2, these estimated sample sizes were developed assuming an extremely large number of sites and of patients on

ART. As countries develop their national protocols, a country-specific sample size can be calculated by using national data on (i) the number of people on ART approximately 48 months prior to survey initiation (for example, if the survey starts in January 2015, this would be the number of people on ART 48 months earlier, that is, as of the end of 2010) and (ii) on the number of clinics where these patients were being

**Table 4: Standard sample sizes calculations for surveillance of viral load suppression among individuals receiving ART for  $\geq 48$  months**

Number of clinics to be sampled	Number of samples per clinic	Total sample size
17	35	595
18	33	594
19	30	570
20	28	560
21	27	567
22	25	550
23	24	552
24	22	528
25	21	525
26	20	520
27	19	513
28	19	532
29	18	522
30	17	510
31	16	496
32	16	512
33	15	495
34	15	510
35	14	490
36	14	504
37	13	481
38	13	494
39	13	507
40	12	480

Notes: assumed prevalence of VL suppression among individuals sampled = 70%, confidence interval width =  $\pm 6\%$ , laboratory failure rate of 15%, proportion of individuals sampled still on first-line ART = 95%; proportion of individuals on first-line ART receiving NNRTI-based regimens = 100%

<sup>1</sup> Adults are generally defined as being 18-years-old and above, though the minimum age may be country specific.

<sup>2</sup> This does not require patients to have been on ART continuously for  $\geq 48$  months. Therefore, for assessing duration on ART, treatment interruptions should be ignored.

treated. This adaptation process will result in a decrease in the estimated total sample size required to reach the desired confidence interval.<sup>1</sup>

The figures presented in Table 3 provide an estimate of standard sample sizes (based on the assumptions discussed in Section 7.2.1) that can be useful for budgetary and planning purposes.

From the tables above, it is evident that as more clinics are sampled, the overall number of patients enrolled from each clinic and the overall sample size required for the survey tends to decrease. This occurs because sampling more clinics reduces the *design effect* by allowing for a more representative sample of clinics (see Statistical Appendix).

### 7.2.3 Countries with few ART clinics

As discussed in Section 6.2.3, countries with few ART sites may choose to sample all clinics. In this case, the standard sample size is 417. The number of sampled patients at each

clinic should be proportional to the size of the clinic (that is, the number of eligible patients at that clinic). For example, if 25% of patients initiate treatment at one clinic in the country, 25% of the actual sample size should be collected from this clinic.

This estimated total sample size was developed assuming an extremely large number of patients on ART. As countries develop their national protocols, a country-specific sample size can be calculated by using national data on the number of people on ART approximately 48 months prior to survey initiation (for example, if the survey starts in January 2015, this would be the number of people on ART as of the end of 2010). This adaptation process will likely result in a decrease in the estimated total sample size required to reach the same confidence interval.

<sup>1</sup> These adjustments can be performed using the Excel-based calculator discussed in section 9.5.

## 8. LABORATORY METHODS

### 8.1 Specimen collection, handling, processing and tracking

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 HIVDR genotyping<sup>1</sup>. DBS specimen should be collected and handled according to the WHO Guidance for DBS specimen collection and handling for HIVDR testing<sup>2</sup>. Countries using plasma specimens for this survey should refer to the WHO recommendations on plasma collection, processing and storage for HIVDR testing.<sup>3</sup>

### 8.2 HIVDR genotyping and quality assurance of sequences

Specimens collected should be tested in WHO-designated HIVDR genotyping laboratories. These laboratories are members of the WHO HIVResNet Laboratory Network,

undergo a rigorous inspection process and participate in annual proficiency panel testing. Use of 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 HIVDR testing, it is encouraged to send specimens to a WHO-designated regional or specialized laboratory. A list of WHO-designated laboratories may be found on the WHO HIVDR webpage.<sup>4</sup>

Designated laboratories perform extensive quality assurance of sequences and follow the WHO Laboratory Standard Operating Procedures (SOP) for Post-Testing Quality Assurance of HIVDR Genotyping. This SOP outlines steps for standardized and automated chromatogram interpretation using Web Recall, quality assurance using MEGA and additional quality assurance and HIVDR interpretation using Stanford HIVdb. This document will be available on the WHO website.

<sup>1</sup> Bertagnolio S, Parkin NT, Jordan M, Brooks J, García-Lerma JG (2010). 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.* 12(4):195-208.

<sup>2</sup> [http://www.who.int/hiv/topics/drugresistance/dbs\\_protocol.pdf](http://www.who.int/hiv/topics/drugresistance/dbs_protocol.pdf)

<sup>3</sup> [http://www.who.int/hiv/pub/drugresistance/hiv\\_reslab\\_strategy.pdf](http://www.who.int/hiv/pub/drugresistance/hiv_reslab_strategy.pdf)

<sup>4</sup> <http://www.who.int/hiv/topics/drugresistance/en/>



## 9. IMPLEMENTATION CONSIDERATIONS

### 9.1 Duration of the survey, patient screening and sampling

To ensure results are available to decision-makers in a timely fashion, it is preferable to limit the duration of patient sampling to a maximum of six months. Once ART clinics to be included in the survey have been selected, a convenient starting date for the survey is chosen. All patients attending the sampled clinics should be screened using the inclusion/exclusion criteria discussed in Sections 6.1 and 7.1. All eligible patients should be consecutively enrolled until the required sample size per clinic is reached or until the maximum enrolment period of six months has passed. Individual clinics can stop sampling patients if the desired sample size is reached earlier.

It is important that all clinics continue to screen and report the numbers of patients receiving ART for 12 ( $\pm 3$ ) and/or  $\geq 48$  months that are observed at the clinic for at least the first three months after the survey start date, even in clinics where the required sample size is reached before then. Information on actual patient accrual is important to develop an accurate measure of relative clinic sizes and will be used at the analysis stage.

For eligible patients, the survey should proceed in two steps:

- **Step 1** (minimum information step): obtain verbal consent and collect minimum information: clinic ID, patient ID, date when ART was initiated for the first time.
- **Step 2** (blood draw step): obtain full consent, collect the remaining necessary information (see Section 8.3 below) and obtain blood draw.

In clinics where the required number of specimens is obtained in less than three months, Step 1 should continue to be applied until the end of the third month.

### 9.2 List of variables to be collected

#### 9.2.1 Patient-level information

##### 9.2.1.1 Patient-level epidemiological and laboratory variables to be captured for all patients who will have blood drawn for VL assessment and genotyping

This section describes the minimal set of patient information that must be captured in the survey database. Some will be obtained using a questionnaire applied to patients at the time of enrolment, while others will come from laboratory records. Once eligible patients have been identified, the following information must be captured for all patients who will have blood drawn for VL assessment and genotyping:

#### Clinical/demographic information

- Clinic ID
- Patient ID (see Box 2 for identification convention)
- Date when ART was initiated for the first time
- Age
- Gender (female, male, other)
- Current ART line (first<sup>1</sup> line /second line/third line/unknown)
- First-line regimen prescribed: list the drugs (if the information is available from medical records)
- Date when second-line therapy was initiated
- Second-line regimen prescribed: list the drugs (if available from medical records)
- Date when third-line ART was initiated
- Third-line regimen prescribed: list the drugs (if available from medical records)

#### Laboratory information

- Specimen ID (see Box 2 for identification convention)
- VL testing successful and results available? (yes/no)
- VL (copies/mL) result from *survey* blood draw
- If VL  $\geq 1000$  copies/mL, reverse transcriptase (RT) region of pol gene successfully<sup>2</sup> sequenced? (yes/no)
- If VL  $\geq 1000$  copies/mL, protease (PR) region of pol gene successfully<sup>3</sup> sequenced? (yes/no/not applicable)
- If VL  $\geq 1000$  copies/mL, INI region of pol gene successfully sequenced? (yes/no/not applicable)
- Drug resistance (see also Section 9). For all drugs, choose the appropriate level according to the Stanford HIV db algorithm interpretation: susceptible, potential low-level, low-level, intermediate, or high-level resistance

##### 9.2.1.2 Minimum information to be captured for all eligible patients for at least the first three months from survey initiation

As discussed in Section 9.1, if the recruitment quota in a particular clinic is reached before the six-month limit, specimen collection can stop at that clinic. However, if this happens within the first three months of survey initiation, these clinics should continue to screen and report the number of patients receiving ART for 12 ( $\pm 3$ ) and/or  $\geq 48$  months observed at the clinic during at least the first three months

<sup>1</sup> The first-line regimen may be the first ART regimen which the patient was prescribed, or it may be an alternative first-line regimen that was started as a substitution. A patient is still classified as being on first line if his/her treatment has been changed from one first-line ART regimen to another first-line ART regimen (intra-class substitution), for example due to adverse events or toxicity.

<sup>2</sup> A specimen is considered to be successfully sequenced only when it passes the appropriate quality assurance as recommended by WHO.

<sup>3</sup> A specimen is considered to be successfully sequenced only when it passes the appropriate quality assurance as recommended by WHO.

## Box 2: Convention for assigning patient identification numbers

Individuals enrolled in the survey will be assigned a survey identification (SID) number, or unique survey ID. This number will be used to identify the patient as well as the sequence generated by the genotyping assay and is composed of the following five elements delimited by a dash character ("-"):

- country abbreviation: the ISO standard 3-letter abbreviation<sup>4</sup>
- survey type: ADR12 (for an ADR survey of 12-month timepoint), ADR48 (for an ADR survey of  $\geq 48$  month timepoint)
- 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

*For example, if the "University HIV Clinic" was a site that participated in a national survey of ADR at 12 months in South Africa in 2014, a participant's ADR12-SID would look like this: ZAF-ADR12-2014-UHC-0001*

*For example, if the "University HIV Clinic" was a site that participated in a national survey of ADR at  $\geq 48$  months in South Africa in 2014, a participant's ADR48-SID would look like this: ZAF-ADR48-2014-UHC-0001*

of the survey period. In this case, the following information should be collected from eligible individuals on ART after the recruitment quota has been reached:

- i. Clinic ID
- ii. Patient ID
- iii. Date when ART was initiated for the first time

### 9.2.2 Clinic-level information

In addition to individual patient-level information, the following information should be collected for each clinic included in the survey:

- i. Clinic name
- ii. Clinic ID
- iii. Date of survey initiation (DD/MM/YYYY)
- iv. Date when specimen collection ended (DD/MM/YYYY)
- v. If specimen collection lasts less than three months, date that last patient was screened (DD/MM/YYYY); if specimen collection lasts more than three months, date when specimen collection ended (DD/MM/YYYY)
- vi. Number of individuals on ART for  $12 \pm 3$  months (and/or  $\geq 48$  months) between date of survey initiation and date when patient screening ended (if specimen collection ends earlier than three months)

- vii. Estimated number of individuals who have been on ART for  $12 \pm 3$  months (and/or  $\geq 48$  months) during a six-month period<sup>2</sup>
- viii. Clinic size as contained in the table used for systematic sampling (an example of systematic sampling table is contained in Annex 1.1)
- ix. If stratification is used, specify stratum name (for example, old/new clinics) to which each clinic belongs
- x. Type of clinic: urban/rural

In countries utilizing methods in Annex 1.6 to develop a nationally representative estimate of retention, the following information should be collected for each clinic included in the survey.

- xi. The total number of eligible patient records for retention review in the clinic
- xii. The number of patient records reviewed in the clinic.
- xiii. The number of patients retained on treatment at 12 months among the reviewed records
- xiv. The number of patients who stopped treatment among the reviewed records.
- xv. The number of patients who died among the reviewed records.

<sup>2</sup> This estimate is achieved by multiplying two numbers. The first number is equal to the number of patients on ART for  $12 \pm 3$  months (or  $\geq 48$  months) observed at that clinic between the date of survey initiation and the date when patient screening ends. The second number is equal to 180 divided by the number of days between the date of survey initiation and the date when patient screening ends. Their product is an estimate of the number of patients on ART for  $12 \pm 3$  months (or  $\geq 48$  months) observed at that clinic during a 6 month period.



- xvi. The number of patients with documented transfer out of the clinic among the reviewed records.
- xvii. The number of patients lost to follow-up among the reviewed records.

### 9.2.3 Survey-level information

- i. Survey type (e.g., ADR12 or ADR48)
- ii. Total number of clinics sampled
- iii. Total number of clinics in the sampling table (an example of systematic sampling table is contained in Annex 1)
- iv. If stratification was used, total number of clinics in each sampling table
- v. Sampling interval from systematic sampling table
- vi. If stratification was used, sampling interval for each stratum

## 9.3 Patient under-enrolment

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.

Prior to the survey initiation, countries should assess whether selected clinics are expected to be able to enrol the required number of patients during the survey period. For example, how likely is it that the required number of people who have been on ART for 12 ( $\pm 3$ ) months or  $\geq 48$  months will be observed at each selected clinic during the survey period?

If countries use the number of people on ART at the end of a previous one year period to establish relative clinic sizes in the sampling table, and assuming a constant distribution of patient visits during the year, the first step is to divide the number of people on ART at each clinic by two, given that the survey is expected to last a maximum of six months. Secondly, a "retention factor" should be applied to estimate how many individuals from the relevant time period are likely to have been retained and could be potentially recruited. For example, when assessing the ability of clinics to enrol patients for the survey of individuals receiving ART for 12 ( $\pm 3$ ) months, the number used to establish clinic size should be multiplied by the estimated retention rate for this group (for example, 85% or 0.85). Similarly, when assessing the ability of clinics to enrol patients for the survey of individuals receiving ART for  $\geq 48$  months, the number used to establish clinic size should be multiplied by the estimate retention rate for this group (for example, 65% or 0.65).

If a country expects to encounter significant under-enrolment (for example, because at random relatively small clinics were sampled), the expected difference should be equally distributed to larger clinics. For example, if the expected under-enrolment is 40 patients, and there are five large clinics<sup>1</sup>, one would sample an additional eight patients from each large clinic.

## 9.4 Repeating the survey

This survey is designed to allow for the assessment of trends of prevalence of viral load suppression in populations receiving ART for 12 ( $\pm 3$ ) and/or  $\geq 48$  months. Thus, it should be repeated periodically, generally every three years or earlier. Countries are advised to update the sampling table and perform a new random sample of clinics to ensure the new survey is adequately representative of changes in the ART programme.

## 9.5 Tools for country adaptation

WHO has developed a user-friendly Excel-based calculator to assist countries to determine their locally appropriate sample sizes based on local information (for example, number of clinics in the country and number of people on ART at the end of a particular period). This calculator can also assist countries to optimize the number of clinics sampled when surveying both timepoints, and develop an integrated sampling plan when conducting Pretreatment HIV drug resistance (PDR) and ADR surveys concomitantly. It will be available for download on WHO's HIVDR website at <http://www.who.int/hiv/topics/drugresistance/en/index.html>.

## 9.6 Combining acquired HIVDR surveys with surveys of pre-treatment HIVDR

WHO has developed a method to survey pretreatment HIVDR in populations initiating ART. Some countries may wish to integrate the pretreatment and ADR surveys by sampling the same clinics for each of the surveys.

As countries explore the appropriateness and feasibility of combining ADR and PDR surveys, the first issue to note is that the clinics sampled for the pretreatment survey will have to be sampled proportionally to the total number of patients on ART by clinic (PPPS sampling). As this design is not the most efficient for the PDR survey, this will affect the PDR sample size. It is also necessary to consider which timepoint will be assessed in the ADR survey:

<sup>1</sup> Definition of "large clinic" is country-specific.

- If only the timepoint will be assessed, then the total number of clinics sampled in the ADR survey will be equal to the number of clinics sampled in the PDR survey, thereby determining the necessary PDR sample size
- If both the 12 ( $\pm 3$ ) months and the  $\geq 48$  month timepoints will be assessed, then a stratified design must be used as described in Section 5. All clinics will be separated into two strata (that is, clinics less than 48 months old, and clinics at least 48 months old). The PDR sample size will be distributed across each stratum according to the proportion of patients attending each type of clinic.
- If only the  $\geq 48$  month timepoint will be assessed, then a stratified design must be used as described above. All clinics will be separated into two strata (that is, clinics less than 48 months old, and clinics at least 48 months old). The PDR sample size will be distributed across each stratum according to the proportion of patients attending each type of clinic. This design will require sampling clinics that are less than 48 months old. These clinics will contribute patients to the PDR survey but not to the ADR  $\geq 48$  months timepoint.

# 10. DATA ANALYSIS

Once all data collection has been completed, prevalence estimated described in Table 1 will be calculated. Data will be weighted taking into account observed clinic-level patient accrual, number of patients screened, and the number of individuals with sequences genotyped. Guidance on data analysis is provided in Annex 1.4. Additional technical background can be found in the Statistical Appendix.

If nationally representative data on retention are available or are obtained through the methods described in Annex 1.6, an estimate of viral load suppression, adjusted for retention, will also be calculated.

For this survey, the Stanford HIVdb algorithm<sup>1</sup> is used to classify HIVDR. The Stanford algorithm classifies HIVDR at five levels: susceptible, potential low-level, low-level, intermediate or high-level drug resistance.

Outcomes 3a and 3b measure prevalence of ANY HIVDR, defined as low-level, intermediate, or high-level resistance according to the Stanford HIVdb to one or more of the following drugs or drug classes: NVP, EFV, any N(t)RTI, DRV/r, LPV/r or ATV/r<sup>2</sup>. Sequences classified as susceptible

and potential low-level resistance are considered as having no HIVDR.

Outcome 3c measures prevalence of HIVDR to NNRTI. Resistance to this class is defined as low-level, intermediate or high-level resistance (according to the Stanford HIVdb) to NVP, EFV, or both. Sequences classified as susceptible and potential low-level resistance are considered as having no HIVDR.

Once all data collection has been completed, prevalence estimates described in Table 1 will be calculated. If nationally representative data on retention are available or are obtained through the methods described in Annex 1.6, an estimate of viral load suppression, adjusted for retention, will also be calculated. Data will be weighted taking into account the number of patients who initiated therapy 12 months prior to survey initiation, observed clinic-level patient accrual, number of patients screened, and the number of individuals with sequences genotyped. Guidance on data analysis is provided in Annex 1.4. Additional technical background can be found in the Statistical Appendix.

<sup>1</sup> Available at: <http://sierra2.stanford.edu/sierra/servlet/JSierra>

<sup>2</sup> Integrase inhibitors should not be included.

# ANNEXES

## Annex 1.1: Selecting the clinics to survey

The section describes how to sample clinics from the list of all ART clinics in the country. Sampling of clinics is performed using systematic sampling to generate probability proportional to proxy size samples (PPPS)<sup>1</sup>.

To execute systematic sampling, all clinics providing ART in the country are listed (Table A1). To enhance the geographical representativeness of the sample, clinics can be listed by administrative or geographical area. Within geographical regions, clinics should be listed in order of size.

Operationally, a) List all eligible clinics providing ART by region and size along with the number of patients on ART at the end of the previous calendar year (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, f) Select clinics based off of the random starting-point, sampling interval and cumulative population size. Table A1 below illustrates these steps in greater detail.

### Detailed instructions

1. List geographical or administrative regions within the country in alphabetical order.
2. Within each region, list all clinics providing ART in that region in order of size (largest to smallest).
3. Record the number of eligible patients who were receiving ART at the end of the previous calendar year by clinic.
4. 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.
5. Determine the sampling interval by dividing the cumulative population size over 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  $13666/20 = 683.3$ , rounded to 683.

6. 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 example, the random number obtained in this example was 500.
7. Select clinics based off the random starting-point, sampling interval and cumulative population size.
  - a. Select the first clinic in which the cumulative size is greater than or equal to the random number. Clinic E has a cumulative population size of 500. Because Clinic E is the first clinic such that 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. 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 our example, Clinic S is selected twice. If a clinic is picked twice, for example, then twice the sample size must be taken from this clinic. For example, if the sample size is 23 per clinic for the first timepoint, 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: Systematic sampling table for clinic selection, survey among people on ART for 12 ( $\pm 3$ ) months**

A	B	C	D	E	F
Region	Clinic name	Number of patients on ART at the end of previous calendar year	Cumulative total of eligible patients	Selection	Sample clinic
A	Clinic A	300	300		
A	Clinic B	111	411		
A	Clinic C	53	464		
A	Clinic D	20	484		

<sup>1</sup> Lohr, Sampling: Design and analysis, 2nd edition, Section 6.2.

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

## Annex 1.2: Guaranteeing representation from all regions

To guarantee that at least one clinic is sampled from each region, a country must determine prior to sampling the minimum number of clinics that must be sampled. The method is as follows:

- Prepare the table that will be used for systematic sampling (one row per clinic) with clinics sorted by region
  - » 1<sup>st</sup> column: region
  - » 2<sup>nd</sup> column: clinic name
  - » 3<sup>rd</sup> column: clinic size
- For each region, add up the sizes of all clinics in the region to determine the region size
- Identify the size of the smallest region
- Determine the preferred number of sampled clinics,  $n$  (for example, 20), based on feasibility
- Calculate the sampling interval,  $SI = (\text{sum of all clinic sizes over all regions})/n$
- Test whether the sampling interval is SMALLER than the SMALLEST region size
  - » If the sampling interval is smaller than the smallest region size, the preferred number of clinics is appropriate and **at least one clinic per region will be sampled with certainty**
  - » If the sampling interval is greater than the smallest region size, increase the preferred number of clinics to sample one additional clinic (for example,  $n$  was 20, now  $n$  is 21), and repeat the last two steps. Continue to increase number of clinics until an appropriate number is identified.

For example, using the data presented in Annex 1.1, if a country wants to guarantee that at least one clinic is sampled from each of the eight regions (A through H), it is necessary to follow the procedure described above to determine the minimum number of clinics to sample.

- For each region, add up the sizes of all clinics in the region to determine the region size

Region	A	B	C	D	E	F	G	H
Size	500	879	2,532	1,905	2,297	1,285	3,315	953

- Identify the smallest region.
  - » A is the smallest region with 500 patients observed.
- Determine the preferred number of sampled clinics,  $n$ , based on feasibility.
  - » For example, the country would like to sample  $n = 20$  clinics.
- Calculate the sampling interval,  $SI$ .
  - » The sampling interval is the sum of all clinic sizes, 13,666, divided by the number of clinics to be sampled, 20.  $SI = 13,666/20 = 683$ .

- Test whether the sampling interval is SMALLER than the SMALLEST region size.
  - » The sampling interval is larger than the size of region A (size = 500). Thus, it is not guaranteed that at least one clinic from region A will be sampled.
- Since the sampling interval is too large to guarantee at least one clinic from region A, increase the number of clinics,  $n$ , until the sampling interval is SMALLER than the SMALLEST region size (that is, less than 500).
  - » For example, for  $n = 27$ ,  $SI = 13,666/27 = 506.1481$ .
  - » For example, for  $n = 28$ ,  $SI = 13,666/28 = 488.0714$ .
  - » Thus,  $n = 28$  is the fewest number of clinics that must be sampled to guarantee at least one clinic from each region is sampled using systematic sampling.
  - » If 28 clinics are too many, the country should consider combining small regions into larger regions.

## Annex 1.3: Stratification

Countries may choose to stratify clinics based on certain characteristics, including clinic location (for example, urban/rural) and clinic type (for example, primary/secondary/tertiary, etc.). The benefit of stratification is that countries can pre-determine how many clinics are sampled from each stratum. If done appropriately, stratification can increase the precision of the survey. If done inappropriately, stratification can decrease the precision of the survey.

Stratification should only be employed if the country (1) has a strong desire to fix the number of clinics sampled per stratum (for example, to guarantee the same number of clinics per region), or (2) wishes to ensure sufficient sample size to make precise stratum-specific statements, or (3) wishes to adjust the design so that different sample sizes are requested of different strata (for example, to define a small clinic stratum and require only a small number of patients per small clinic), or (4) has knowledge of a clinic-level factor that is associated with viral load suppression, retention, or HIVDR (for example, urban clinics tend to have higher VLS than rural clinics).

The following guidelines should be observed for the appropriate implementation of stratification:

### Defining strata

- The number of stratifying variables should be limited to only those that are most associated with the outcomes or those that are most relevant to the survey designers. Extraneous stratifying variables should be avoided.
- The number of levels of the stratifying variables should be limited, where possible. For example, urban/rural is two levels; primary/secondary/tertiary is three levels.
- If more than one stratifying variable is used, the number of strata is equal to the product of the number of levels in each stratifying variable. For example, if both urban/rural and primary/secondary/tertiary are used, there are six ( $= 2 \times 3$ ) levels. Again, the number of combinations should be limited.



- If a regional variable is used, larger regions (such as north, central and south) are preferable over smaller regions (such as districts, of which there may be many). Small regions (as defined by having few eligible patients) may be combined with other similar regions.
- All strata must be non-empty. If there are no urban tertiary clinics, this stratum should be eliminated. If there are very few patients in urban tertiary clinics, this stratum should be combined with a similar stratum, such as urban secondary clinics.

In general, no single stratum should be too small, that is, contain too few eligible patients.

## Designing the survey

### Step 1: calculate the effective sample size

The first step in designing the survey, after determining the number of stratifying variables, is to calculate the effective sample size. The method for calculating the effective sample size given an assumed prevalence, a desired precision, a desired number of clinics to be sampled, and the predetermined number of strata, is described in the Statistical Appendix.

- » EXAMPLE: for the early timepoint, the prevalence of VLS is 85%, confidence interval half-width is 5%, the desired number of clinics sampled is 20, and there are two strata: urban and rural. Then, the total effective sample size is 226 (without adjustment for laboratory failure and design effect).

### Step 2: allocate the effective sample size to different strata

Next, the effective sample is allocated to the strata. In general, the effective sample size should be allocated proportionally to each stratum.

- » EXAMPLE: If 60% of eligible patients reside in urban areas and 40% in rural areas,  $226 \times 0.60 = 136$  is the effective sample size in the urban stratum, and  $226 \times 0.40 = 90$  is the effective sample size in the rural stratum.

Any design that does not allocate the sample size proportionally to the size of the strata will be less efficient (such as taking equally sized samples from regions of vastly different size) and therefore not recommended.

### Step 3: calculate the actual sample size

Now, in each stratum, the effective sample is used to determine an appropriate actual sample size. The effective sample size must be inflated by the design effect, estimated laboratory failure, and other factors described for sample size calculations.

- » EXAMPLE: assuming a desired total number of clinics of 20, the country could choose to sample 12 urban clinics with 22 patients per clinic, and eight rural clinics with 22 patients per clinic (adjusting for laboratory failure, expected proportion of patients

on first-line regimens, and expected proportion of patients on NNRTI-based first-line regimens)

Countries do not need to sample the same number per clinic across all strata.

- » EXAMPLE: Because urban clinics are much larger than rural clinics, the country could choose a different design. The country could choose to sample 8 urban clinics with 34 patients per clinic, and 12 rural clinics with 14 patients per clinic.
- » COUNTRIES MUST SAMPLE AT LEAST TWO CLINICS FROM EACH STRATUM, even if the stratum is small. This is very important for the analysis stage of the survey.

## Executing the survey

To perform the survey, countries create a sampling table for each stratum. Within each stratum, countries can use systematic sampling to sample the desired number of clinics. The general method for systematic sampling is the same as described previously in Annex 1.1.

- » EXAMPLE: List all urban clinics. Use systematic sampling to sample 12 urban clinics (if first design described in Step 3 is selected). List all rural clinics. Use systematic sampling to sample 8 rural clinics.

## Annex 1.4: Data analysis plan

User-friendly instructions are provided below for data analysis in Stata.

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. Stata, SUDAAN and R's survey package allow users to specify finite population corrections at each stage of sampling. SAS and WesVar do not allow users to specify finite population corrections beyond the first stage of sampling. Epi Info does not allow users to specify any finite population corrections.

Deviations in Stata code for survey designs using **Stratification** or sampling **All Sites** are indicated by these key words. A survey design without stratification and only a subset of sites is referred to as a **Standard** design. Surveys can collect data of patients receiving ART for 12 ( $\pm 3$ ) months, (operationally defined as 9-15 months), defined as the **Early Timepoint**. Countries conducting the survey of the **Early Timepoint** may also collect data on Retention either via a **Retention Census** or a Retention Survey (see section 4.4 and Annex 1.6).

We provide below a worked out example of a survey that assesses only the outcomes of the **12 ( $\pm 3$ ) months Timepoint** and has a standard design, that is, a survey that samples a subset of clinics and does not use stratification. In this example, **Retention** is assessed via a **Retention Survey**. In this example, 19 unique clinics are selected via PPPS systematic sampling from the sample population described in Annex 1.1. From each clinic, 23 patients are enrolled into the VLS survey, and 83 patient records are assessed for retention. Because Clinic "S" was sampled twice during systematic sampling, 46 patients and 166 patient records are included from this clinic. Clinic-specific sample sizes also vary because of differential laboratory failure or under-enrollment.

For purposes of clarity and to ensure standardization of participant identification and specimen labeling (Box 2), in this example the Early Timepoint is referred to as ADR12 and the Late Timepoint is referred to as ADR48.

## I. Summarize site-level information

In Excel, create a spreadsheet summarizing the necessary information from each site.

1. List the unique site IDs of the sampled sites in a column labelled "SITE\_ID"<sup>1,2</sup>.
2. **If Stratification is used:** List the stratum IDs of each of the sampled sites in a column labelled "STRATUM\_ID".
3. Calculate the site sampling weight for each site and record it in a column labelled "SITE\_SAMPLING\_WT".
  - a. **Standard:** The site sampling weight is equal to the sampling interval from the systematic sampling table (for example, 683) divided by the estimated site size from Column C of the systematic sampling table (for example, for site E, the weight is  $683/16 = 42.6875$ , rounded to 42.688). Note: smaller sites will have larger site sampling weights.
  - b. **Stratification:** For a site from a particular stratum, the site sampling weight is equal to the sampling interval from the stratum-specific systematic sampling table divided by the estimated site size from the same stratum-specific systematic sampling table. Note: each stratum will have a different sampling interval.
  - c. **All sites:** The site sampling weight is equal to 1 for all sites.
4. **Early Timepoint (ADR12):** Summarize the data collected for the 12 ( $\pm 3$ ) months timepoint.
  - a. List the estimated 6-month eligible population sizes in a column labelled "N\_ADR12\_PATIENTS\_6MOS". The eligible population for the early timepoint is defined as the number of patients on treatment for 9-15 months during the 6-month survey period at each site (this value may be extrapolated using the method described in Section 9.2.2 if the eligible population size is observed for less than 6 months).
  - b. List the number of 9-15 months patients with amplified VL results at each site in a column labelled "N\_ADR12\_SPECIMENS\_AMPLIFIED".
  - c. List the number of 9-15 months patients who are virally suppressed at each site in a column labelled "N\_ADR12\_SPECIMENS\_SUPPRESSED".
  - d. List the number of 9-15 months patients who are not virally suppressed at each site in a column labelled "N\_ADR12\_SPECIMENS\_NOT\_SUPPRESSED".
  - e. List the number of 9-15 months patients with genotyped sequences at each site in a column labelled "N\_ADR12\_SPECIMENS\_GENOTYPED".
5. **Late Timepoint (ADR48):** Summarize the data collected for the late timepoint.
  - a. List the estimated 6-month eligible population sizes in a column labelled "N\_ADR48\_PATIENTS\_6MOS". The eligible population is defined as patients on treatment for  $\geq 48$  months during the 6-month survey period at each site (this value may be extrapolated using the method described in Section 9.2.2 if the eligible population size is observed for less than 6 months).
  - b. List the number of  $\geq 48$  month patients with amplified VL results at each site in a column labelled "N\_ADR48\_SPECIMENS\_AMPLIFIED".
  - c. List the number of  $\geq 48$  month patients who are virally suppressed at each site in a column labelled "N\_ADR48\_SPECIMENS\_SUPPRESSED".
  - d. List the number of  $\geq 48$  month patients who are not virally suppressed at each site in a column labelled "N\_ADR48\_SPECIMENS\_NOT\_SUPPRESSED".
  - e. List the number of  $\geq 48$  month patients with genotyped sequences at each site in a column labelled "N\_ADR48\_SPECIMENS\_GENOTYPED".
6. **Retention:** In this example, retention is calculated through information collected at the survey sites following the method described in Annex 1.6:
  - a. List the number of patients eligible for the retention record review by site in a column labelled "N\_RETENTION\_ELIGIBLE".
  - b. List the number of patients whose records were reviewed for retention by site in a column labelled "N\_REVIEWED".
  - c. List the number of patients whose records were reviewed for retention by site, excluding those with documented transfer out of the site, in a column labelled "N\_REVIEWED\_EXCL\_TRANSFERRERED".
  - d. List the number of patients whose records indicate that they were retained at 12 months by site in a column labelled "N\_RETAINED".
7. Save data in a spreadsheet, such as "ADR\_SITE\_DATA.xlsx". Table A2 above provides an example of site-level data for a standard design with only the 12 ( $\pm 3$ ) month timepoint assessed.

<sup>1</sup> At the analysis stage, all variable names should be indicated with capital letters without quotation marks in the column headers.

<sup>2</sup> It is recommended that the site ID correspond exactly to the three-letter site code described in Box 2. In this highly simplified example, a one- to two-letter site code is used.



**Table A2: Example of site-level data for standard design with only Early Timepoint assessed**

SITE_ID	SITE_SAMPLING_WT	N_ADR12_PATIENTS_6MOS	N_ADR12_SPECIMENS_AMPLIFIED	N_ADR12_SPECIMENS_SUPPRESSED	N_ADR12_SPECIMENS_NOT_SUPPRESSED	N_ADR12_SPECIMENS_GENOTYPED	N_RETENTION_ELIGIBLE	N_REVIEWED	N_REVIEWED_EXCL_TRANSFERRED	N_RETAINED
E	42.688	18	17	16	1	1	18	18	17	15
G	1.935	402	23	20	3	3	220	83	77	62
J	1.131	580	21	16	5	4	252	83	77	64
K	1.138	633	22	19	3	3	274	83	78	69
M	1.783	420	21	18	3	3	168	83	78	63
S	0.994	788	40	36	4	4	317	166	156	133
...										

## II. Summarize patient-level viral load suppression information

In Excel, create a spreadsheet summarizing the necessary information for each patient sampled for the viral load suppression portion of the survey.

- List the unique patient ID in a column labelled "ID".
- List the site ID in a column labelled "SITE\_ID" (must be identical to the site ID in the column labelled "SITE\_ID" in the site-level spreadsheet).
- Early Timepoint (ADR12):** List a binary variable indicating whether the patient is on therapy for 9-15 months in a column labelled "ADR12\_PATIENT\_BN".
- Late Timepoint (ADR48):** List a binary variable indicating whether the patient is on therapy for  $\geq 48$  months in a column labelled "ADR48\_PATIENT\_BN".
- List a binary variable indicating whether a patient is on first-line ART in a column labelled "FIRST\_LINE\_BN" (1 if on first line; 0 if not on first line; missing if line not recorded).
- List a binary variable indicating whether a patient is on an NNRTI-based first-line ART regimen in a column labelled "FIRST\_LINE\_NNRTI\_BN" (1 if amplified; 0 if specimen was collected but no data are available; missing if no specimen was collected).
- List a binary variable indicating whether a patient's VL specimen was amplified in a column labelled "AMPLIFIED\_BN" (1 if amplified; 0 if specimen was collected but no data are available; missing if no specimen was collected).
- List a binary variable indicating whether a patient was virally suppressed in a column labelled "VL\_SUPPRESSED\_BN" (1 if suppressed; 0 if not suppressed; missing if no specimen was collected or specimen was collected but no data are available).
- For patients who were not virally suppressed, list a binary variable indicating whether a patient's specimen was genotyped in a column labelled "GENOTYPED\_BN" (1 if genotyped; 0 if specimen was eligible for genotyping but results are not available; missing if specimen was not eligible for genotyping).

- For patients whose samples were successfully genotyped, list a binary variable indicating HIVDR in a column labelled "ANY\_HIVDR\_BN" (1 if HIVDR detected; 0 if no HIVDR detected; missing if data are not available, specimen was not eligible for genotyping, or no specimen was collected).
- Save data in a spreadsheet, such as "ADR\_PT\_DATA\_VLS.xlsx".

Table A3 below provides an example of patient-level data for viral load suppression at the early timepoint.

In the above sample table:

- Patient ADR12-2014-E-0001 (from Site E) was on a first-line NNRTI-based regimen for 9-15 months, was not virally suppressed, and had evidence of HIVDR.
- Patient ADR12-2014-E-0002 (from Site E) was not on a first-line regimen (though on treatment for 19-15 months was not virally suppressed, and did not have a genotype available).
- Patient ADR12-2014-E-0003 (from Site E) was on an unknown regimen for 9-15 months and was virally suppressed.
- Patient ADR12-2014-G-0001 (from Site G) was on a first-line non-NNRTI-based regimen for 9-15 months, and did not have an amplified viral load result available.

## III. Import viral load suppression data into Stata

- Import site data using the import data option (*File/Import/Excel Spreadsheet*)<sup>1</sup>. Use the Browse button to identify the spreadsheet. Select the option to import the first row as variable names. Change the variable case to upper to preserve variable names.
- Save site data as a .dta file using the save option (*File/Save*). In this example, we save the data as "ADR\_SITE\_DATA.dta".
- Import patient viral load suppression data using the import data option (*File/Import/Excel Spreadsheet*). Use the Browse button to identify the spreadsheet. Select the

<sup>1</sup> (Menu/Option/Sub-option) indicates using the drop-down menus to select an option.

Table A3: Example of patient-level data for Viral Load Suppression at 12 ( $\pm 3$ ) months Timepoint

ID	SITE_ID	ADR12_PATIENT_BN	FIRST_LINE_BN	FIRST_LINE_NNRTI_BN	AMPLIFIED_BN	VL_SUPPRESSED_BN	GENOTYPED_BN	ANY_HIVDR_BN
ADR12-2014-E-0001*	E	1	1	1	1	0	1	1
ADR12-2014-E-0002*	E	1	0	0	1	0	0	
ADR12-2014-E-0003*	E	1			1	1		
....								
ADR12-2014-G-0001*	G	1	1	0	0			

\*simplified patient IDs, but true patient IDs should follow naming conventions in Box 2.

option to import the first row as variable names. Change the variable case to upper to preserve variable names.

4. Save patient viral load suppression data as a .dta file using the save option (*File/Save*). In this example, we save the data as "ADR\_PATIENT\_DATA\_VLS.dta". Press "Yes" to overwrite data currently in memory.
5. Merge the two datasets using the merge option (*Data/Combine datasets/Merge two datasets*). Select the "Many-to-one" option. Select or type in "SITE\_ID" as the Key Variable. Use the Browse button to select "ADR\_SITE\_DATA.dta". Press OK.
6. **Early Timepoint:** Keep only observations that are eligible for the Early Timepoint and have amplified VL results. In the command line, type in "keep if ADR12\_PATIENT\_BN==1 & AMPLIFIED\_BN==1"<sup>1</sup>.
7. **Late Timepoint:** Keep only observations that are eligible for the early timepoint and have amplified VL results. In the command line, type in "keep if ADR48\_PATIENT\_BN==1 & AMPLIFIED\_BN==1".

**Both Early Timepoint and Late Timepoint:** Run commands for Early Timepoint only (skip Step 7). Repeat process for Late Timepoint only (skip Step 6).

#### IV. Create survey weights and other necessary variables for analysing Viral Load Suppression and HIVDR

The directions below are for analysing the **Early Timepoint**, but they can be readily extended to analysis of the **Late Timepoint** by substituting all EARLY variables for their LATE equivalents. .

1. Create survey weight for Outcomes 1a, 1b and 1c. In the command line, type in "generate OUTCOME1\_WT = SITE\_SAMPLING\_WT\*(N\_ADR12\_PATIENTS\_6MOS/N\_ADR12\_SPECIMENS\_AMPLIFIED)".
2. Create survey weight for Outcomes 3a, 3b, 3c and 4. In the command line, type in the following:
  - a. "generate OUTCOME3\_WT = OUTCOME1\_WT".
  - b. "replace OUTCOME3\_WT = OUTCOME1\_WT\*(N\_ADR12\_SPECIMENS\_NOT\_SUPPRESSED/N\_ADR12\_SPECIMENS\_GENOTYPED) if GENOTYPED\_BN==1".

3. Create a variable indicating the total number of sites in the sampling frame (prior to systematic sampling).
  - a. **Standard or All Site:** In our example, there were 41 total sites. In the command line, type in "generate N\_TOTAL\_SITES = 41". In practice, replace 41 with your country-specific number.
  - b. **Stratification:** The variable should refer to the number of total sites in the stratum-specific sampling frame. In the command line, type in "generate N\_TOTAL\_SITES = .". Then, for each stratum, use the replace command to identify the stratum-specific number. For example, if there are 50 sites in stratum 1 and 40 sites in stratum 2, type in "replace N\_TOTAL\_SITES = 50 if STRATUM\_ID == 1" and "replace N\_TOTAL\_SITES = 40 if STRATUM\_ID == 2".
4. Create a variable to be used for reporting results for global aggregation. In the command line, type in "generate POP\_SIZE = 1".

#### V. Declare survey design and analyse data for Viral Load Suppression and HIVDR

1. Declare survey design for Outcomes 1a, 1b and 1c (*Statistics/Survey data analysis/Setup & utilities/Declare survey design for dataset*).
  - a. In the Main tab, change "Number of stages" to 2.
  - b. Select "Site\_ID" as the "Stage 1: Sampling units".
  - c. **Standard or All Sites:** Leave "Stage 1: Strata" blank.
  - d. **Stratification:** Select "STRATUM\_ID" as the "Stage 1: Strata".
  - e. Select "N\_TOTAL\_SITES" as the "Stage 1: Finite pop. correction".
  - f. Select "ID" as the "Stage 2: Sampling units".
  - g. Leave "Stage 2: Strata" blank.
  - h. Select "N\_ADR12\_PATIENTS\_6MOS" as the "Stage 2: Finite pop. correction".
  - i. In the Weights tab, select "OUTCOME1\_WT" as the "Sampling weight variable".
  - j. In the SE tab, select "Center at the grand mean" for Strata with a single sampling unit. Press OK.
2. Analyse Outcome 1a. (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*).
  - a. In the Model tab, select "VL\_SUPPRESSED\_BN" as the "Variable". Press OK.

<sup>1</sup> For all Stata commands in the command line, do not include the quotation marks. Type only the text between the quotations marks.

- b. The point estimate, standard error, and 95% confidence interval for the prevalence of viral load suppression are located in the row labelled "1".
- c. In the output, the number of PSUs (Primary Sampling Units) is equal to the number of unique clinics sampled in the survey. The design degrees of freedom are equal to the number of PSUs minus the number of strata, where the number of strata is equal to 1 if no stratification is used.

#### . svy linearized : proportion VL\_SUPPRESSED\_BN

Survey: Proportion estimation				
Number of strata =	1	Number of obs =	417	
Number of PSUs =	19	Population size =	13408.1	
		Design df =	18	
	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
VL_SUPPRESSED_BN				
0	.1647105	.0167436	.1324727	.2029581
1	.8352895	.0167436	.7970419	.8675273

3. Analyse Outcome 1b. (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*).
  - a. In the Model tab, select "VL\_SUPPRESSED\_BN" as the "Variable(s)".
  - b. In the if/in/over tab, type in "FIRST\_LINE\_BN==1" (selects only patients on first-line regimens) into the "If: (expression)" box. Press OK.
  - c. The point estimate, standard error, and 95% confidence interval for the prevalence of viral load suppression among patients on first-line regimens are located in the row labelled "1".
  - d. The number of observations used to calculate this outcome is labelled "Subpop. no. obs". The value labelled "Number of obs" is not a meaningful quantity and should be disregarded.

#### . svy linearized, subpop(if FIRST\_LINE\_BN==1) : proportion VL\_SUPPRESSED\_BN

Survey: Proportion estimation				
Number of strata =	1	Number of obs =	417	
Number of PSUs =	19	Population size =	13408.1	
		Subpop. no. obs =	311	
		Subpop. size =	9924.89	
		Design df =	18	
	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
VL_SUPPRESSED_BN				
0	.1712323	.0199764	.1332345	.2173494
1	.8287677	.0199764	.7826506	.8667655

4. Analyse Outcome 1c in the same manner as Outcome 1b, selecting patients on first-line NNRTI-based regimens as the appropriate subpopulation ("FIRST\_LINE\_NNRTI\_BN==1").
5. Declare survey design for Outcomes 3a, 3b and 4 (*Statistics/Survey data analysis/Setup & utilities/Declare survey design for dataset*).
  - a. In the Main and SE tabs, select the same options described in Step 1.
  - b. In the Weights tab, select "OUTCOME3\_WT" as the "Sampling weight variable". Press OK.
6. Analyse Outcome 3a. (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*).
  - a. In the Model tab, select "ANY\_HIVDR\_BN" as the "Variable".
  - b. In the if/in/over tab, type in "VL\_SUPPRESSED\_BN==0" (selects only patients with viral load failure) into the "If: (expression)" box. Press OK.
  - c. The point estimate, standard error, and 95% confidence interval for the prevalence of HIVDR among patients with VL failure are located in the row labelled "1".

#### . svy linearized, subpop(if VL\_SUPPRESSED\_BN==0) : proportion ANY\_HIVDR\_BN

Survey: Proportion estimation				
Number of strata =	1	Number of obs =	409	
Number of PSUs =	19	Population size =	13408.1	
		Subpop. no. obs =	61	
		Subpop. size =	2208.46	
		Design df =	18	
	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
ANY_HIVDR_BN				
0	.3703671	.0641972	.248055	.5119297
1	.6296329	.0641972	.4880703	.751945

7. Analyse Outcome 3b. (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*).
  - a. In the Model tab, select "ANY\_HIVDR\_BN" as the "Variable".
  - b. In the if/in/over tab, type in "VL\_SUPPRESSED\_BN==0 & FIRST\_LINE\_BN==1" (selects only patients with viral load failure AND on first-line regimens) into the "If: (expression)" box. Press OK.
  - c. The point estimate, standard error, and 95% confidence interval for the prevalence of HIVDR among patients with VL failure and on first-line regimens are located in the row labelled "1".

**. svy linearized, subpop(if VL\_SUPPRESSED\_BN==0 & FIRST\_LINE\_BN==1) : proportion ANY\_HIVDR\_BN**

Survey: Proportion estimation			
Number of strata =	1	Number of obs =	412
Number of PSUs =	19	Population size =	13505.9
		Subpop. no. obs =	49
		Subpop. size =	1737.59
		Design df =	18

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
ANY_HIVDR_BN				
0	.3620066	.0710998	.2290949	.5200138
1	.6379934	.0710998	.4799862	.7709051

8. Analyse Outcome 3c in the same manner as Outcome 3b, selecting patients on first-line NNRTI-based regimens who are not virally suppressed as the appropriate subpopulation ("VL\_SUPPRESSED\_BN==0 & FIRST\_LINE\_NNRTI\_BN==1").
9. Analyse Outcome 4. (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*).
  - a. In the Model tab, select "ANY\_HIVDR\_BN" as the "Variable". Press OK.
  - b. The point estimate, standard error, and 95% confidence interval for the prevalence of HIVDR are located in the row labelled "1".

**. svy linearized : proportion ANY\_HIVDR\_BN**

Survey: Proportion estimation			
Number of strata =	1	Number of obs =	409
Number of PSUs =	19	Population size =	13408.1
		Design df =	18

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
ANY_HIVDR_BN				
0	.8962929	.0153352	.8593722	.9243733
1	.1037071	.0153352	.0756267	.1406278

10. Analyse data to report for aggregated global analysis. Example shown is the prevalence of HIV drug resistance among patients with VL failure, restricted to patients not on first-line regimens.
  - a. Repeat Step 5 to declare survey design used for Outcomes 3a, 3b and 3c.
  - b. Repeat Step 6, but in the if/in/over tab, type in "VL\_SUPPRESSED\_BN==0 & FIRST\_LINE\_BN==0" (selects only patients with VL failure who are not on first-line regimens) into the "If: (expression)" box. Press OK.

The point estimate and standard error for the prevalence of viral load suppression among patients not on first-line regimens are located in the row labelled "1". The number of observations used to construct this estimate is labelled "Subpop. no. obs"; this is equal to the number of patients not on first-line regimens with viral load results available. The design degrees of freedom are labelled "Design df".

**. svy linearized, subpop(if VL\_SUPPRESSED\_BN==0 & FIRST\_LINE\_BN==0) : proportion ANY\_HIVDR\_BN**

Survey: Proportion estimation			
Number of strata =	1	Number of obs =	414
Number of PSUs =	19	Population size =	13568
		Subpop. no. obs =	12
		Subpop. size =	470.874
		Design df =	18

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
ANY_HIVDR_BN				
0	.4012185	.1706397	.1309441	.7487311
1	.5987815	.1706397	.2512689	.8690559

- c. To aggregate the data at a global level, we also need the numerator of the HIVDR prevalence estimate (and its associated standard error) and the denominator of the HIVDR prevalence estimate (and its associated standard error). In this scenario, the numerator is an estimate of the total number of patients on treatment for 9-15 months in the country during the 6-month survey period who are not virally suppressed, have detected HIVDR, and are not on first-line regimens. The denominator is an estimate of the total number of patients on treatment for 9-15 months in the country during the 6-month survey period who are not virally suppressed and are not on first-line regimens. The prevalence is equal to the numerator divided by the denominator. Select (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Totals*). In the Variables box in the Model tab, type or select "ANY\_HIVDR\_BN POP\_SIZE". In the if/in/over tab, type in "VL\_SUPPRESSED\_BN==0 & FIRST\_LINE\_BN==0" (selects only patients with VL failure who are not on first-line regimens) into the "If: (expression)" box. Press OK.
- d. The numerator estimate and its standard error are located in the row labeled "ANY\_HIVDR\_BN". The denominator estimate and its standard error are located in the row labeled "POP\_SIZE".



**. svy linearized, subpop(if VL\_SUPPRESSED\_BN==0 & FIRST\_LINE\_BN==0) : total ANY\_HIVDR\_BN POP\_SIZE**

Survey: Total estimation			
Number of strata =	1	Number of obs =	414
Number of PSUs =	19	Population size =	13568
		Subpop. no. obs =	12
		Subpop. size =	470.874
		Design df =	18

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
ANY_HIVDR_BN	281.9504	109.4564	51.99112	511.9097
POP_SIZE	470.8736	147.0261	161.9833	779.764

## VI. Reformat retention data for analysis

- Open site-level Stata dataset labeled "ADR\_SITE\_DATA.dta". If necessary, press "Yes" to overwrite data currently in memory.
- Type the following series of commands. After entering these three commands, the reformatted dataset will have one line per retention record reviewed, excluding documented transfers. Each record will have a binary variable which is equal to 1 if the reviewed patient is retained on treatment, and 0 if the patient is not retained on treatment.
  - "expand 2, gen(RETAINED\_BN)"
  - "expand N\_RETAINED if RETAINED\_BN==1"
  - "expand N\_REVIEWED\_EXCL\_TRANSFERRED - N\_RETAINED if RETAINED\_BN==0"
- Save patient retention dataset as "ADR12RETENTION\_DATA.dta".

## VII. Create survey weights and other necessary variables for analysing retention

- Create survey weight for Outcome 2a (retention). In the command line, type in "generate OUTCOME2A\_WT = SITE\_SAMPLING\_WT\*(N\_RETENTION\_ELIGIBLE/N\_REVIEWED)".
- Create a variable indicating the total number of sites in the sampling frame (prior to systematic sampling).
  - Standard or All Sites:** In our example, there were 41 total sites in the sampling frame (systematic sampling table). In the command line, type in "generate N\_TOTAL\_SITES = 41". In practice, replace 41 with your country-specific number.
  - Stratification:** The variable should refer to the number of total sites in the stratum-specific sampling frame. In the command line, type in "generate N\_TOTAL\_SITES = ". Then, for each stratum, use the replace command to identify the stratum-specific number. For example, if there are 50 sites in stratum 1 and 40 sites in stratum 2, type in "replace N\_TOTAL\_

SITES = 50 if STRATUM\_ID == 1" and "replace N\_TOTAL\_SITES = 40 if STRATUM\_ID == 2".

- Create a variable to be used for reporting results for global aggregation. In the command line, type in "generate POP\_SIZE = 1".

## VIII. Declare survey design and analyse data for Outcome 2a (retention)

- Declare survey design for Outcome 2a (retention) (*Statistics/Survey data analysis/Setup & utilities/Declare survey design for dataset*).
  - In the Main tab, change "Number of stages" to 2.
  - Select "SITE\_ID" as the "Stage 1: Sampling units".
  - Standard or All Sites:** Leave "Stage 1: Strata" blank.
  - Stratification:** Select "STRATUM\_ID" as the "Stage 1: Strata".
  - Select "N\_TOTAL\_SITES" as the "Stage 1: Finite pop. correction".
  - Type in "\_n" as the "Stage 2: Sampling units".
  - Leave "Stage 2: Strata" blank.
  - Select "N\_RETENTION\_ELIGIBLE" as the "Stage 2: Finite pop. correction".
  - In the Weights tab, select "OUTCOME2A\_WT" as the "Sampling weight variable".
  - In the SE tab, select "Center at the grand mean" for Strata with a single sampling unit. Press OK.
- Analyse Outcome 2a. (*Statistics/Survey data analysis/Means, proportions, ratios, totals/Proportions*).
  - In the Model tab, select "RETAINED\_BN" as the "Variable". Press OK.
  - The point estimate, standard error, and 95% confidence interval for the prevalence of retention are located in the row labelled "1".

**. svy linearized : proportion RETAINED\_BN (running proportion on estimation sample)**

Survey: Proportion estimation			
Number of strata =	1	Number of obs =	1463
Number of PSUs =	19	Population size =	7195.7
		Design df =	18

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
RETAINED_BN				
0	.1659568	.009645	.146671	.1872221
1	.8340432	.009645	.8127779	.853329

## IX. Analyse Outcome 2b – Adjusted viral load suppression

1. Open (*File/Open*) site-level data "ADR\_SITE\_DATA.dta". If prompted, press yes to replace data currently in memory.
2. Create a variable indicating the total number of sites in the sampling frame (prior to systematic sampling).
  - a. **Standard or All Sites:** In our example, there were 41 total sites. In the command line, type in "generate N\_TOTAL\_SITES = 41". In practice, replace 41 with your country-specific number.
  - b. **Stratification:** The variable should refer to the number of total sites in the stratum-specific sampling frame. In the command line, type in "generate N\_TOTAL\_SITES = .". Then, for each stratum, use the replace command to identify the stratum-specific number. For example, if there are 50 sites in stratum 1 and 40 sites in stratum 2, type in "replace N\_TOTAL\_SITES = 50 if STRATUM\_ID == 1" and "replace N\_TOTAL\_SITES = 40 if STRATUM\_ID == 2".
3. Generate a new variable that is the site-specific estimate of the prevalence of viral load suppression among retained patients. In the command line, type in "generate PREVALENCE\_VLS = N\_ADR12\_SPECIMENS\_SUPPRESSED/N\_ADR12\_SPECIMENS\_AMPLIFIED".
4. Generate a new variable that is the site-specific estimate of the prevalence of retention. In the command line, type in "generate PREVALENCE\_RET = N\_RETAINED/N\_REVIEWED\_EXCL\_TRANSFERRED".
5. Generate a new variable that is the site-specific estimate of the adjusted prevalence of viral load suppression. In the command line, type in "generate PREVALENCE\_ADJ = PREVALENCE\_VLS\*PREVALENCE\_RET".
6. Generate a new variable summarizing the number of sites sampled.
  - a. **Standard or All Sites:** In the command line, type in "generate N\_SITES\_SAMPLED = \_N".
  - b. **Stratification:** In the command line, type in "bysort STRATUM\_ID : generate N\_SITES\_SAMPLED = \_N".
7. Generate a new variable that is the sampling weight for Outcome 2b. In the command line, type in "generate OUTCOME2B\_WT = SITE\_SAMPLING\_WT\*N\_RETENTION\_ELIGIBLE".
8. Declare survey design for Outcome 2b (*Statistics/Survey data analysis/Setup & utilities/Declare survey design for dataset*).
  - a. In the Main tab, let "Number of stages" equal 1.
  - b. Select "SITE\_ID" as the "Stage 1: Sampling units".
  - c. **Standard or All Sites:** Leave "Stage 1: Strata" blank.
  - d. **Stratification:** Select "STRATUM\_ID" as the "Stage 1: Strata".
  - e. Select "N\_TOTAL\_SITES" as the "Stage 1: Finite pop. correction".
9. Run the following commands in the command line:
  - a. "svy linearized : total PREVALENCE\_ADJ".
  - b. "matrix B\_MAT = e(b)".
  - c. "scalar PREVALENCE\_ADJ\_PT = B\_MAT[1,1]".
  - d. "scalar DF = e(df\_r)".
  - e. "scalar DENOM = e(N\_pop)".
  - f. "matrix R\_MAT = r(table)".
  - g. "scalar VAR\_TERM1 = R\_MAT[2,1]^2".
  - h. "gen VAR\_VLS = (1-N\_ADR12\_SPECIMENS\_AMPLIFIED/N\_ADR12\_Y\_PATIENTS\_6MOS)\*PREVALENCE\_VLS\*(1-PREVALENCE\_VLS)/N\_ADR12\_SPECIMENS\_AMPLIFIED".
  - i. "gen VAR\_RET = (1-N\_REVIEWED\_EXCL\_TRANSFERRED/N\_RETENTION\_ELIGIBLE)\*PREVALENCE\_RET\*(1-PREVALENCE\_RET)/N\_REVIEWED\_EXCL\_TRANSFERRED".
  - j. "gen VAR\_TERM2\_SITE = (1/DENOM^2)\*N\_SITES\_SAMPLED/N\_TOTAL\_SITES\*(OUTCOME2B\_WT)^2\*(PREVALENCE\_VLS^2\*VAR\_RET+PREVALENCE\_RET^2\*VAR\_VLS-VAR\_VLS\*VAR\_RET)".
  - k. "summarize VAR\_TERM2\_SITE".
  - l. "scalar VAR\_TERM2 = r(sum)".
  - m. "scalar PREVALENCE\_ADJ\_SE = sqrt(VAR\_TERM1 + VAR\_TERM2)".
  - n. "scalar PREVALENCE\_ADJ\_LOGIT\_LB = ln(PREVALENCE\_ADJ\_PT/(1-PREVALENCE\_ADJ\_PT))-invttail(DF,0.025)\*PREVALENCE\_ADJ\_SE/(PREVALENCE\_ADJ\_PT\*(1-PREVALENCE\_ADJ\_PT))".
  - o. "scalar PREVALENCE\_ADJ\_LOGIT\_UB = ln(PREVALENCE\_ADJ\_PT/(1-PREVALENCE\_ADJ\_PT))+invttail(DF,0.025)\*PREVALENCE\_ADJ\_SE/(PREVALENCE\_ADJ\_PT\*(1-PREVALENCE\_ADJ\_PT))".
  - p. "scalar PREVALENCE\_ADJ\_LB = exp(PREVALENCE\_ADJ\_LOGIT\_LB)/(1+exp(PREVALENCE\_ADJ\_LOGIT\_LB))".
  - q. "scalar PREVALENCE\_ADJ\_UB = exp(PREVALENCE\_ADJ\_LOGIT\_UB)/(1+exp(PREVALENCE\_ADJ\_LOGIT\_UB))".
  - r. "scalar drop VAR\_TERM1 VAR\_TERM2 PREVALENCE\_ADJ\_LOGIT\_LB PREVALENCE\_ADJ\_LOGIT\_UB".
  - s. "scalar list".
  - t. "PREVALENCE\_ADJ\_PT" is the point estimate of Outcome 2b, with standard error equal to "PREVALENCE\_ADJ\_SE" and 95% confidence interval ("PREVALENCE\_ADJ\_LB", "PREVALENCE\_ADJ\_UB").
- f. In the Weights tab, select "OUTCOME2B\_WT" as the "Sampling weight variable".
- g. In the SE tab, select "Center at the grand mean" for Strata with a single sampling unit. Press OK

### . scalar list

PREVALENCE_ADJ_UB =	.72971392
PREVALENCE_ADJ_LB =	.6690616
PREVALENCE_ADJ_SE =	.01444775
DENOM =	7651.753
DF =	18
PREVALENCE_ADJ_PT =	.70026451

Point estimate = 70.02% with 95% confidence interval (66.91%, 72.97%).

## Annex 1.5: Reporting of HIVDR data

All countries are expected to report to WHO a dataset including (1) individual patient information (demographic and matching laboratory data), (2) clinic data and (3) survey variables discussed in Section 9.2, in addition to the patient sequences in FASTA file format. It is recommended that sequences ID, patients ID and specimens ID numbers be identical.

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

For this survey, the Stanford HIVdb algorithm is used to classify HIVDR. The Stanford HIVdb algorithm classifies HIVDR at five levels: susceptible, potential low-level, low-level, intermediate or high-level resistance. Sequences classified as susceptible and potential low-level resistance are considered as having “no HIVDR”.

The utilization of these different categories is summarized below:

### 1. HIVDR by individual drug

When reporting HIVDR for all drugs individually (except ATV/r, LPV/r or DRV/r), sequences classified as low-, intermediate-, or high-level resistance according to the Stanford HIVdb are classified as “HIV drug resistance”.

For ATV/r, LPV/r or DRV/r, the level of resistance should be classified as “HIV drug resistance” (if either high or intermediate level according to the Stanford HIVdb) or “low-level HIV drug resistance” (if low-level resistance according to the Stanford HIVdb). Sequences classified as susceptible

and potential low-level resistance are considered as having “no HIVDR”.

### 2. HIVDR by drug class

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

- NNRTI class includes any NNRTI
- NRTI class includes any NRTI
- boosted PI class includes only DRV/r, or LPV/r or ATV/r
- Integrase inhibitor class includes any integrase inhibitor

Sequences classified as low-, intermediate- or high-level resistance according to the Stanford HIVdb are aggregated as “HIV drug resistance”. This also applies to boosted PI’s.

### 3. Any HIVDR

“Any HIV drug resistance” is defined in sequences classified as low-, intermediate-, or high-level resistance according to the Stanford HIVdb with respect to one or more of the following drugs: NVP, EFV, any N(t)RTI, DRV/r, LPV/r or ATV/r<sup>1</sup>.

## Annex 1.6: Calculating a nationally representative estimate of retention

It is possible to calculate a nationally representative estimate of retention at 12 months by reviewing a predetermined number of existing patient records from the same clinics sampled for the ADR survey.

The first step is to identify the relevant cohort of patients whose records will be sampled. This should be done with two considerations in mind:

- The cohort should cover a period of 12 months to avoid potential biases in the estimation of retention associated with seasonal fluctuations.
- A patient is assumed to be lost to follow-up only if he/she has not returned to the clinic for three months. Therefore, the relevant cohort should be defined in a way that allows for sufficient time for the last patient who initiated ART in the cohort to be assessed as being lost to follow-up or not.

Ideally, the cohort of patients eligible for the retention record review would be identical to the cohort of patients eligible for VL suppression and HIVDR assessment. If this approach were adopted, the retention assessment would need to be delayed by as many as 12 months after survey initiation to allow patients to achieve 12 months on ART plus three months of additional follow-up. To ensure that the retention assessment can be initiated on the same date as the ADR survey, a different cohort is identified that initiated ART 15

<sup>1</sup> INI should not be included.

months (12 months plus three months) to 27 months (24 months plus three months) prior to the survey start date. Operationally, this means that, if the survey start date is, for example, 01 August 2014, the retention assessment should review records of patients who initiated ART between 01 May 2012 (that is, 27 months prior to the survey start date) and 30 April 2013 (that is, 15 months prior to the survey start date). In this way, a patient who initiated ART on 30 April 2013 would have until 31 July 2014 to have his status ascertained. Thus, the retention assessment can be initiated immediately alongside the survey on August 01 2014. One limitation of using this cohort is that patients eligible for retention record review did not initiate ART during the same time period as patients eligible for viral load suppression and HIVDR assessment; in fact, retention patients initiated ART prior to patients participating in the ADR survey. However, it is assumed that retention in care does not drastically change between the two cohorts.

- Once the cohort of patients is identified, a predefined number of patient files are randomly sampled for review. The number of patient files to be reviewed is the same for all clinics and it is based, among other things, on the desired precision of the national retention estimate (for example, 5%). An Excel-based tool has been developed to help countries determine the total number of patient files to be reviewed (see Section 9.5). Clinics with insufficient patient records should sample all available eligible records. Random sampling can be performed consecutively if the number of files is small, or systematically (for example, every 5th file) if the number of files is large.
- Information on whether these patients were retained or officially transferred-out at 12 months after ART start date should be recorded. Documented transfers-out should be censored from numerator and denominator.
- It is assumed that patients who have not been retained are classified as having virological failure.

- The national retention estimate is calculated as the overall proportion of patients who are alive and on ART among those who initiated ART during the relevant 12-month cohort period.
- As the clinics are randomly sampled, this measure of retention is nationally representative.

The sample size required for estimating retention is affected by two key variables:

1. the expected patient retention rate 12 months after therapy initiation, assumed to be 85%
2. the desired confidence interval half-width, default set to be  $\pm 5\%$ .

Table A4 provides combined standard sample size calculations for estimating Outcome 1a for the early timepoint and retention, as well the resulting confidence interval for Outcome 2b. Using this method, the resulting confidence interval for Outcome 2b (adjusted prevalence of viral load suppression) will be approximately 6%.

In countries sampling all ART clinics the total sample size to estimate Outcome 1a is 364, and the number of patient files to be reviewed to estimate retention is 310. The resulting confidence interval of the retention-adjusted prevalence of viral load suppression will be approximately 5.9%.



**Table A4: Standard sample size calculations for surveillance of individuals receiving ART for 12 ( $\pm$ 3) months**

Number of clinics to be sampled	Outcome 1a		Retention		Outcome 2b
	Number of samples per clinic to estimate Outcome 1a	Total sample size for Outcome 1a	Number of files to be reviewed per clinic to estimate retention	Total number of files to be reviewed to estimate retention	Resulting CI $\pm$ of Outcome 2b
17	28	476	512	8704	6.0%
18	26	468	187	3366	6.1%
19	24	456	115	2185	6.0%
20	23	460	83	1660	6.1%
21	21	441	65	1365	6.1%
22	20	440	53	1166	6.0%
23	19	437	45	1035	6.0%
24	18	432	39	936	6.0%
25	18	450	35	875	6.1%
26	17	442	31	806	6.0%
27	16	432	28	756	6.0%
28	15	420	26	728	6.0%
29	15	435	24	696	6.0%
30	14	420	22	660	6.0%
31	14	434	21	651	6.0%
32	13	416	19	608	6.0%
33	13	429	18	594	6.1%
34	12	408	17	578	6.1%
35	12	420	16	560	6.0%
36	12	432	16	576	5.9%
37	11	407	15	555	6.0%
38	11	418	14	532	6.0%
39	11	429	14	546	5.9%
40	10	400	13	520	5.9%

Notes: assumed prevalence of VL suppression among individuals sampled = 85%, confidence interval half-width for Outcome 1a =  $\pm$ 5%, laboratory failure rate of 15%, proportion of individuals sampled receiving first-line ART = 95%.

# STATISTICAL APPENDIX

## 1. Sample size calculations: Prevalence of viral load suppression

To determine the necessary sample size for the survey, we start by determining the effective sample size for estimating the prevalence of viral load suppression. The effective sample size refers to the number of patients,  $k_{eff}$ , we would need to sample to achieve a desired confidence interval half-width if we were conducting a simple random sample. The effective sample size is determined by the prevalence of the outcome and the desired width of the confidence interval. The effective sample size is then multiplied by the design effect to yield the actual sample size of the survey.

To determine the effective sample size, consider the following formula for a Wald-type confidence interval. Here,  $\tilde{p}_{VLS}$  refers to the prevalence of viral load suppression among individuals sampled,  $n$  refers to the number of clinics sampled, and  $df$  are the design degrees of freedom:

$$95\% \text{ CI} = \left( \tilde{p}_{VLS} - t_{df,0.975} \sqrt{\frac{\tilde{p}_{VLS}(1 - \tilde{p}_{VLS})}{k_{eff}}}, \tilde{p}_{VLS} + t_{df,0.975} \sqrt{\frac{\tilde{p}_{VLS}(1 - \tilde{p}_{VLS})}{k_{eff}}} \right)$$

The design degrees of freedom are defined as  $df = (\text{number of clinics sampled}) - (\text{number of strata})$ . If stratification is not used,  $df = n - 1$ .

The half-width of this confidence interval (referring to the distance from the midpoint to either end of the confidence interval) is:

$$L = t_{df,0.975} \sqrt{\frac{\tilde{p}_{VLS}(1 - \tilde{p}_{VLS})}{k_{eff}}}$$

So, the effective sample size  $k_{eff}$  is the smallest sample size such that  $t_{df,0.975} \sqrt{\frac{\tilde{p}_{VLS}(1 - \tilde{p}_{VLS})}{k_{eff}}}$  is less than  $L$ . Once the effective sample size is calculated, it has to be adjusted to reflect the design effect. This is addressed in Section 2 below.

The effective sample size can be calculated using the following formula<sup>1</sup>:  $k_{eff} = \frac{(t_{df,0.975})^2 \tilde{p}_{VLS}(1 - \tilde{p}_{VLS})}{L^2}$

The effective sample sizes for the 12 and  $\geq 48$  month timepoints are calculated in Box A1.

Note: Because the method for calculating a confidence interval in the setting of clustered surveys uses a t distribution with degrees of freedom equal to the design degrees of freedom, our effective sample size is also a function of the number of clinics sampled. When the design degrees of freedom are large (40 or greater), it is standard to use the normal approximation ( $z$ ) instead of the  $t$ :  $z_{0.975} \approx t_{df,0.975}$

This simplifies calculations, but introduces an error, of relative size less than 3%. Since this design requires sampling of around 17–40 clinics, the design degrees of freedom will be small, and it is thus inadvisable to make this simplification. At 16 degrees of freedom, for example, the relative error is almost 8%. The consequence of using this simplification would be an underestimation of the total sample size required to achieve a given confidence interval half-width.

The effective sample size must be inflated to determine the actual sample size required for the survey. The amount by which the sample size is increased is called the design effect. The elements of the study design that contribute to the design effect are (1) clustering of the outcome by clinic ( $DEFF_{clust}$ ), and (2) imperfect information from using data from a previous year or from a slightly different population ( $DEFF_{info}$ ). These elements are described in greater detail below.

## Box A1: Calculating the effective sample size for the 12 ( $\pm 3$ ) and $\geq 48$ month timepoints

For the 12-month timepoint, if a country plans to sample  $n = 20$  clinics without stratification ( $df = 20 - 1 = 19$ , and  $t_{19,0.975} = 2.09$ ), assumes that the prevalence of viral load suppression is 85%, ( $\tilde{p}_{VLS} = 0.8$ ), and desires a confidence interval width of  $\pm 5\%$  ( $L = 0.05$ ), then the following effective sample size is required:

$$k_{eff} = \frac{2.093^2 \times 0.85 \times (1 - 0.85)}{0.05^2} = 223.4$$

Thus, the required effective sample size in this example is approximately 224 individuals. Thus, if a simple random sample of patients on treatment for 12 ( $\pm 3$ ) months in the country were being performed, a sample size of 224 individuals would be required to achieve the desired precision.

For the  $\geq 48$ -month timepoint, if a country plans to sample  $n = 20$  clinics without stratification ( $df = 20 - 1 = 19$ , and  $t_{19,0.975} = 2.09$ ), assumes that the prevalence of viral load suppression is 70%, ( $\tilde{p}_{VLS} = 0.70$ ), and desires a confidence interval width of  $\pm 5\%$  ( $L = 0.05$ ), then the following effective sample size is required:

$$k_{eff} = \frac{2.093^2 \times 0.70 \times (1 - 0.70)}{0.05^2} = 367.9$$

Thus, the required effective sample size in this example is approximately 368 individuals. Thus, if a simple random sample of patients on treatment for  $\geq 48$  months in the country were being performed, a sample size of 368 individuals would be required to achieve the desired precision.

In general, the effective sample size increases as the desired confidence interval width decreases and as the estimated prevalence approaches 50%.

## 2. Calculating the contribution to the design effect due to clustering of the outcome by clinic

It is first necessary to calculate the design effect due to clustering of the outcome. Clustering of the outcome occurs because the amount of viral suppression varies by clinic. Patients from the same clinic may have more similar viral load outcomes than patients from different clinics in the same country. The similarity of patients within clinics is measured via the intraclass correlation coefficient, or ICC.

If  $m$  is the number of patients sampled per clinic and  $ICC_{VLS}$  is the estimated intraclass correlation for the viral load suppression outcome, the design effect due to clustering can be estimated using the following formula:

$$DEFF_{clust} = 1 + (m - 1)ICC_{VLS}$$

The design effect due to clustering increases as more patients are sampled from the same clinics ( $m$  increases).

In order to estimate the intraclass correlation, global data from WHO's Global HIV Drug Resistance Report 2012 were used. For each clinic in each country, the estimated probability of viral load suppression was calculated at the 12-month timepoint after censoring patients with documented transfer to another clinic. Intraclass correlation was then estimated using an analysis of variance estimator<sup>1</sup>. Although intraclass correlation is defined as capturing the clustering of outcomes by clinics within the same country, clinics in the dataset were not separated by country.

Using the raw data, with observed prevalence of viral load suppression of 89% at 12 ( $\pm 3$ ) months after treatment initiation, the estimated ICC was very low ( $ICC_{VLS,raw} = 0.0032$ ). As the assumed prevalence of viral load suppression for the 12-month timepoint was 85%, and since the ICC and prevalence are generally correlated, the ICC was adjusted to reflect the assumed prevalence. To perform this adjustment, a linear model predicting  $\log(ICC)$  by  $\log(\text{prevalence})$  was applied. The equation used:  $ICC_{VLS} = \exp(0.91 \times \ln[(1 - \tilde{p}_{VLS}) / (1 - 0.891)]) \times ICC_{VLS,raw}$ .<sup>2</sup> For the assumed prevalence of 85% at the

<sup>1</sup> Ridout et al. (1999) Estimating intraclass correlation for binary data. *Biometrics* 55, 137-148.

<sup>2</sup> Guillford et al. (2005) Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *Journal of Clinical Epidemiology* 58, 246-251. Note: log prevalence coefficient used is 0.91 (from HTA data)

12-month timepoint, the multiplicative factor was 1.34, resulting in an estimated  $ICC_{VLS,12} = 1.34 \times 0.0032 \approx 0.004$ . As the assumed prevalence of viral load suppression for the  $\geq 48$  month timepoint was 70%, a similar adjustment was performed; the multiplicative factor was 2.51, resulting in an estimated ICC of  $ICC_{VLS,\geq 48} = 2.51 \times 0.0032 \approx 0.008$ .

It is important to note that there are limitations to these estimates. First of all, the ICC estimates are based on only the data available in the WHO's Global HIV Drug Resistance Report 2012. A 95% confidence interval can be constructed for  $ICC_{VLS,raw}$  using Searle's method<sup>1</sup>, and the resulting interval extends from -0.001425 to 0.01339; thus, the interval is very wide, reflecting the uncertainty in the estimate. Furthermore, the ICC, which is calculated from patients on treatment for 12 months, is more reasonably generalized to the 12-month outcome than the  $\geq 48$ -month outcome. One might posit that the degree of clustering is higher for patients receiving treatment in the same sites for a longer period of time. Thus, as surveys are implemented, it is important that the data obtained be used to better inform the estimate of ICC for future iterations of the survey.

### 3. Calculating the contribution to the design effect due to imperfect weighting information

Ideally, we use probability proportional to size sampling to sample clinics proportional to the number of individuals who have been on ART for 12 ( $\pm 3$ ) or  $\geq 48$  months at each clinic. However, these numbers are generally not available in most countries. Thus, it is recommended that countries employ *probability proportional to proxy size (PPPS) sampling*. In PPPS sampling, clinics are sampled with probabilities proportional to a proxy measure, such as the total number of individuals on ART by clinic. Ideally, the number of patients on treatment for 12 ( $\pm 3$ ) and/or  $\geq 48$  months at each clinic will be highly correlated with the proxy measure; thus, the overall design will be close to proportional. If the number of patients on treatment for 12 ( $\pm 3$ ) months and/or  $\geq 48$  months at each site is not highly correlated with the proxy measure, then the design will be further from proportional and, thus, less efficient. Because information from a previous time period or from a slightly different population is used to conduct the sampling, the weights will not be perfectly proportional. It is expected that the sizes of the eligible populations within the clinics will change over time, although it is assumed that changes in the relative clinic sizes will not be sizable. If clinic populations change dramatically over time, for example because of decentralization, this information should be incorporated into the estimated population sizes used for sampling. The goal is to use estimated population sizes that will be most predictive of the population sizes to be observed during the survey period.

To estimate the effect of imperfect information on the design effect, we use a formula estimating the variance contribution for disproportionate weights<sup>2</sup>. The design effect can be approximated by  $DEFF_{info} = 1 + cv^2(\text{weights})$ , where  $cv()$  refers to the coefficient of variation and *weights* are the survey weights. For PPPS sampling, it is estimated that  $DEFF_{info} = 1.50$ . This corresponds to inflating the sample size by 50% to account for the imperfect information. This number was calculated from observing the differences in population sizes between treatment initiators and patients on ART at clinic in an African country over a two-year period. The value reflects the correlation between a particular cohort in a clinic and the entire clinic population, which is relevant when the population of interest is patients on therapy for 12( $\pm 3$ ) and/or  $\geq 48$  months.

These numbers are approximations, and the true values may also be very country specific. As surveys are implemented, it is recommended that these values be re-evaluated and adjusted as necessary for future iterations of the survey.

### 4. Other sources of design effect

The design effect is also influenced by other sources of variability. For example, different clinics will have different levels of viral load amplification failure. This will induce additional variability in the weights. It is estimated that this source of design effect will be small, so it is ignored in the calculations to simplify the design.

### 5. Calculating the sample size

The design effect for the viral load suppression outcome is estimated using the following equation<sup>3</sup>:

$$DEFF = DEFF_{clust} \times DEFF_{info}$$

For each timepoint, the following procedure can be used to identify an appropriate sample size for the survey:

1 Searle SR. Linear models. New York, Wiley, 1971.

2 Kalton, et al. (2005), Section B, Chapter VI: Estimating components of design effects for use in sample design. In: Household sample surveys in developing and transition countries (p. 110, eqn. 23). New York, United Nations. Retrieved at <http://unstats.un.org/unsd/hhsurveys/>.

3 Park and Lee (2004) Design effects for the weighted mean and total estimators under complex surveys sampling, Survey Methodology 30: 183-193.

- Calculate the necessary effective sample size  $k_{eff}$  for a given number of clinics  $n$
- Determine the appropriate value of  $DEFF_{info}$  based on the sampling design (1.5 for PPS)
- Determine the appropriate value of the ICC, for example  $ICC_{VLS} = 0.004$  for the 12-month timepoint
- Solve the following equation for  $m$ , which is the number of patients to be sampled per clinic for a particular timepoint:

$$m = \frac{1 - ICC_{VLS}}{\left[ \frac{n}{DEFF_{info} k_{eff}} - ICC_{VLS} \right]}$$

- If such an  $m$  does not exist, or if the calculated value of  $m$  is too large to be practical in a particular setting, consider increasing the number of clinics sampled,  $n$ . Because of the design effect due to clustering, sampling a larger number of clinics should require fewer samples per clinic (within rounding error), and it will also require a smaller overall sample size.

The sample size needs to be adjusted for three additional parameters:

1. Laboratory failure; for example, if we expect a 15% laboratory failure rate, we need to divide the required sample size by 0.85 to compensate for only getting 85% of the sample size on average.
2. Expected proportion of patients sampled receiving a first-line regimen. In order to retain statistical power at the analysis stage when considering only patients on first-line regimens, the sample size needs to be adjusted for the expected proportion of patients sampled receiving a first-line regimen. For the sake of simplicity, it is assumed that 95% of patients sampled will be receiving a first-line regimen.
3. Expected proportion of patients sampled on first-line ART receiving NNRTI-based regimens, in this case assumed to be 100%. This should be the last step in the sample size calculations.

$$m_{samp} = m / (0.85 \times 0.95 \times 1)$$

**Table A5: Summary of assumptions for sample size calculations**

	12 ( $\pm 3$ ) months	$\geq 48$ months
Estimated intracluster correlation	0.004	0.008
Expected prevalence of viral load suppression	85%	70%
Expected lab failure rate	15%	15%
Effect of imperfect information for PPS sampling	1.5	1.5
Expected proportion of patients sampled receiving a first-line regimen	95%	95%
Expected proportion of patients sampled on receiving first-line NNRTI-based regimens	100%	100%

## 6. Incorporating the finite population correction

The formula for the design effect due to clustering can be revised to incorporate the predicted effect of the finite population corrections which will be applied at the analysis stage. The design effect due to clustering in the absence of finite population corrections is  $DEFF_{clust} = 1 + (m - 1)ICC_{VLS}$  where  $m$  is the number of patients sampled per clinic and ICC is the intracluster correlation. For a country with  $N$  total clinics in the sampling frame and an average of  $M$  eligible patients per clinic, it can be shown that the design effect due to clustering can be approximated by:

$$DEFF_{clust} \approx \left(1 - \frac{m}{M}\right) + \left[\left(1 - \frac{n}{N}\right)m - \left(1 - \frac{m}{M}\right)\right] ICC_{VLS}$$

The average number of eligible patients per clinic can be estimated as the total number of eligible patients (estimated from available data) divided by the total number of clinics in the sampling frame ( $\bar{M} = M/N$ ). It can be shown that the necessary number of patients per clinic to be sampled per clinic to achieve a desired precision is approximately:

$$m = \frac{1 - ICC_{VLS}}{\frac{n}{k_{eff} DEFF_{info}} - ICC_{VLS} \left(1 - \frac{n}{N}\right) + \frac{N}{M} (1 - ICC_{VLS})}$$

The sample size must then be adjusted for expected viral amplification failure and the expected proportion of patients on first-line therapy. This calculation assumes that at least  $m$  patients can be sampled from all clinics; if fewer than  $m$  patients are sampled from a clinic, the confidence interval will be slightly wider than planned for.

## 7. Sample size calculations when all clinics are sampled

If all clinics in the sampling frame will be included in the survey, the following modifications can be made to the sample size calculations. Briefly, the survey effective sample size is calculated, and this effective sample size is multiplied by a design effect due to imperfect information, the expected laboratory failure, and the expected proportion of patients on first-line regimens. It is not necessary to multiply the calculations by a design effect due to clustering because all clinics in the sampling frame are included in the survey.

The effective sample size necessary to achieve a confidence interval of half-width  $L$  is:  $k_{eff} = \frac{3.84 \tilde{p}_{VLS}(1 - \tilde{p}_{VLS})}{L^2}$

If the finite population correction is incorporated into the calculations (where  $M$  is the total eligible population size in the country), then the effective sample size can be calculated using the following equation:

$$k_{eff} = \frac{M \times 3.84 \tilde{p}_{VLS}(1 - \tilde{p}_{VLS})}{L^2 \times M + 3.84 \tilde{p}_{VLS}(1 - \tilde{p}_{VLS})}$$

Because information on patient enrolment from a prior time period will be used to allocate the sample, it is recommended that the sample size be inflated slightly to account for imperfect information. This is equivalent to adjusting for a design effect for disproportionate weighting. For PPPS sampling, the sample should be inflated by  $DEF_{info} = 1.50$ . Next, the sample size should be inflated by the expected laboratory success rate (85%), the expected proportion of patients on first-line therapy (95%), and the expected proportion of patients sampled on first-line therapy who are receiving NNRTI-based regimens (100%). Thus, the actual sample size is:

$$k_{act} = \frac{k_{eff} \times 1.50}{0.85 \times 0.95} \approx 1.86k_{eff}$$

The actual sample size is then allocated to the clinics proportional to the number of eligible patients expected to be observed during the survey period. For each clinic, the sample size of that clinic is equal to the total sample size,  $k_{act}$ , multiplied by the expected patient accrual at that clinic divided by the expected patient accrual for all clinics included in the survey. For example, if 25% of patients in a country attend a particular clinic, 25% of the sample size should be allocated to that clinic. The per-clinic sample sizes are rounded to the nearest whole number.

## 8. Sample size calculations for nationally representative estimate of retention

The same procedure described for the viral load suppression outcome can be used to calculate necessary sample sizes to achieve a particular confidence interval width for the estimated retention at 12 months. The following parameters should be used: based on data from WHO's Global Update on HIV Treatment 2013<sup>1</sup>, the estimated prevalence of retention at 12 months is assumed to be 85%, that is,  $\tilde{p}_{RET} = 0.85$ . Using data published in WHO's 2012 Global HIV Drug Resistance Report, the estimated intracluster correlation coefficient is  $ICC_{RET,raw} = 0.0713$  with an observed prevalence of 12 month retention of 76.6%. Using the formula described previously to calculate the intracluster correlation,  $ICC_{RET} = \exp(0.91 \times \ln[(1 - \tilde{p}_{RET})/(1 - 0.766)]) \times ICC_{RET,raw}$ <sup>2</sup>. For the assumed prevalence of 85% at the 12-month timepoint, the multiplicative factor was 0.667, resulting in an estimated  $ICC_{RET} = 0.667 \times 0.0713 \approx 0.0475$ . The assumed  $DEF_{info} = 1.5$  because PPPS sampling is used.

As described previously for the viral load suppression outcome, the steps for determining the necessary sample size for the 12 month retention outcome are as follows:

1. Calculate the effective sample size using the assumed prevalence, the desired confidence interval width, and the number of clinics to be sampled.
2. Estimate the intracluster correlation coefficient.
3. Solve for the number of patients to be sampled per clinic for the retention outcome.
4. Adjust the sample size calculations for the expected prevalence of documented transfer, assumed to be 5%, since these patients will be censored from the calculations. Thus, the sample size should be divided by  $(1 - 0.05) = 0.95$ . If desired, the finite population correction can be incorporated using the formulae described above. The total eligible population size is an estimate of the number of patients who initiated therapy over a predefined period 12 months prior to survey initiation (see Annex 1.6).

1 Global update on HIV treatment 2013: Results, impact and opportunities. Geneva, World Health Organization, 2013.

2 Guillford et al. (2005) Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. Journal of Clinical Epidemiology 58, 246-251. Note: log prevalence coefficient used is 0.91 (from HTA data).



## 9. Estimated confidence interval width for Outcome 2b

Given a particular sample size for the viral load measure at 12 months, and given a particular sample size for the retention measure at 12 months, the predicted variance and confidence interval width for the adjusted viral load suppression outcome (Outcome 2b) can be calculated. Let  $m$  be the number of patients sampled per clinic for the 12 month viral load suppression measure (excluding amplification failures), let  $s$  be the number of patients per clinic for the retention measure (excluding documented transfers), and let  $M$  be the total number of patients who initiated therapy 12 months prior to survey initiation.

The estimated confidence interval width without the use of the finite population correction can be calculated using the following equations, where  $n$  is the number of clinics sampled,  $m$  is the number of patients sampled per clinic for the estimation of viral load suppression, and  $s$  is the number of records sampled per clinic for the estimation of retention:

$$\begin{aligned} \text{var}(\hat{p}_{ADJ}) \approx & DEFF_{info} \frac{1}{n} \left\{ \left[ ICC_{VLS} + \frac{1}{m} (1 - ICC_{VLS}) \right] \tilde{p}_{RET}^2 \tilde{p}_{VLS} (1 - \tilde{p}_{VLS}) \right. \\ & + \left[ ICC_{RET} + \frac{1}{s} (1 - ICC_{RET}) \right] \tilde{p}_{VLS}^2 \tilde{p}_{RET} (1 - \tilde{p}_{RET}) \\ & \left. + \left[ ICC_{VLS} + \frac{1}{m} (1 - ICC_{VLS}) \right] \left[ ICC_{RET} + \frac{1}{s} (1 - ICC_{RET}) \right] \tilde{p}_{VLS} (1 - \tilde{p}_{VLS}) \tilde{p}_{RET} (1 - \tilde{p}_{RET}) \right\} \end{aligned}$$

$$CI \text{ HalfWidth} \approx t_{n-1, 0.975} \sqrt{\text{var}(\hat{p}_{ADJ})}$$

The estimated confidence interval width incorporating the finite population correction can be calculated using the following equations, where  $M$  is the estimated total number of patients retained on therapy for 9-15 months and  $S$  is the estimated total number of eligible records in the country for the estimation of retention:

$$\begin{aligned} \text{var}(\hat{p}_{ADJ}) \approx & DEFF_{info} \frac{1}{n} \left\{ \left[ ICC_{VLS} + \left( \frac{1}{m} - \frac{N}{M} \right) (1 - ICC_{VLS}) \right] \tilde{p}_{RET}^2 \tilde{p}_{VLS} (1 - \tilde{p}_{VLS}) \right. \\ & + \left[ ICC_{RET} + \left( \frac{1}{s} - \frac{N}{S} \right) (1 - ICC_{RET}) \right] \tilde{p}_{VLS}^2 \tilde{p}_{RET} (1 - \tilde{p}_{RET}) \\ & + \left[ ICC_{VLS} + \left( \frac{1}{m} - \frac{N}{M} \right) (1 - ICC_{VLS}) \right] \left[ ICC_{RET} + \left( \frac{1}{s} - \frac{N}{S} \right) (1 - ICC_{RET}) \right] \\ & \left. \times \tilde{p}_{VLS} (1 - \tilde{p}_{VLS}) \tilde{p}_{RET} (1 - \tilde{p}_{RET}) \right\} \end{aligned}$$

$$CI \text{ HalfWidth} \approx t_{n-1, 0.975} \sqrt{\text{var}(\hat{p}_{ADJ})}$$

## 10. Data analysis

### 10.1 Data analysis: Clinic sampling weight and clinic size

Once an appropriate design is identified, clinics will be sampled using PPS systematic sampling. In PPS systematic sampling, clinics are sampled proportional to the proxy measure of clinic size. For clinic  $i$ , the proxy measure of clinic size is denoted as  $\tilde{M}_i$ . If the predetermined number of clinics to be selected is  $n^*$ , the probability that a clinic is selected is equal to  $n^* \tilde{M}_i$  divided by the total size of all clinics in the sampling frame,  $\tilde{M} = \sum_{j=1}^N \tilde{M}_j$ .

Note: large clinics may be sampled more than once using this methodology. If a large clinic is sampled twice, this clinic should sample twice as many patients, and so on. In this case, the number of unique clinics selected, denoted  $n$ , is fewer than  $n^*$ . In this setting, it is necessary to distinguish between  $n^*$  and  $n$  in the calculations.

Thus, the clinic sampling weight is equal to the following (where SI is the sampling interval defined in Annex 1.1):

$$w_{clinic, i} = \left( \frac{\tilde{M}}{n^* \tilde{M}_i} \right) = \left( \frac{SI}{\tilde{M}_i} \right)$$

If all clinics are included in the survey, the clinic sampling weight is equal to 1 for all clinics. If a stratified survey is conducted, clinic weights should be constructed separately for each stratum.

For the 12-month timepoint,  $M_i$  is a count of the number of patients attending clinic  $i$  observed during the 6-month survey period that have been on treatment for 9–15 months. For clinics that reach their enrolment quotas before six months, they should continue to count eligible patients for a minimum of three months, and  $M_i$  can be estimated as two times the number of eligible patients attending clinic  $i$  observed during the three-month period. This quantity must also be recorded for the  $\geq 48$ -month timepoint.

## 10.2 Data analysis: Outcome 1a

Outcome 1a measures population-level viral load suppression (VL < 1000 copies/mL) among individuals who have been on ART for 12 (or  $\geq 48$ ) months and who have been retained on treatment. Outcome 1a, therefore, is not adjusted to take into account the proportion of people who no longer attend clinics because they have been lost to care, have died or have stopped treatment. Data analysis for this outcome and all additional outcomes is to be conducted in Stata using the SVY utilities<sup>1</sup>. Even if Stata is not used to conduct the analysis, the Stata SVY manual section on Variance Estimation contains all necessary formulae for calculating the prevalence, variance and 95% confidence interval of each outcome.

The clinic sampling weight is defined in Section 10.1. The patient sampling weight for clinic  $i$  is defined as  $M_i$  divided by the number of patients on treatment for 12 (or  $\geq 48$ ) months with amplified VL data available from that clinic. The overall weight is the product of the clinic and patient sampling weights. All patients with available VL measurements are defined as either having VL suppression (binary variable for VL suppression = 1) or no VL suppression (binary variable for VL suppression = 0). The prevalence is estimated using a ratio. The numerator is an estimate of the total number of patients in the country retained on treatment for 12 (or  $\geq 48$ ) months achieving viral load suppression. The denominator is an estimate of the total number of patients in the country retained on treatment for 12 (or  $\geq 48$ ) months. The variance is calculated using linearization. A 95% confidence interval can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

## 10.3 Data analysis: Outcome 1b

Outcome 1b measures population-level viral load suppression (VL < 1000 copies/mL) among individuals on first-line regimens who have been on ART for 12 (or  $\geq 48$ ) months and who have been retained on treatment. Outcome 1b is a subpopulation analysis of the data used for Outcome 1a. Data analysis is conducted using the same sampling weights described for Outcome 1a. The population is restricted to patients on first-line regimens using the subpopulation command in Stata. All patients on first-line regimens with available VL measurements are defined as either having VL suppression (binary variable for VL suppression = 1) or no VL suppression (binary variable for VL suppression = 0). The prevalence is estimated using a ratio. The numerator is an estimate of the total number of patients in the country retained on first-line regimens for 12 (or  $\geq 48$ ) months achieving viral load suppression. The denominator is an estimate of the total number of patients in the country retained on first-line regimens for 12 (or  $\geq 48$ ) months. The variance is calculated using linearization. A 95% confidence interval can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

## 10.4 Data analysis: Outcome 1c

Outcome 1c measures population-level viral load suppression (VL < 1000 copies/mL) among individuals on NNRTI-based first-line regimens who have been on ART for 12 (or  $\geq 48$ ) months and who have been retained on treatment. Outcome 1c is a subpopulation analysis of the data used for Outcome 1a. Data analysis is conducted using the same sampling weights described for Outcome 1a. The population is restricted to patients on NNRTI-based first-line regimens using the subpopulation command in Stata. All patients on NNRTI-based first-line regimens with available VL measurements are defined as either having VL suppression (binary variable for VL suppression = 1) or no viral load suppression (binary variable for VL suppression = 0). The prevalence is estimated using a ratio. The numerator is an estimate of the total number of patients in the country retained on NNRTI-based first-line regimens for 12 (or  $\geq 48$ ) months achieving viral load suppression. The denominator is an estimate of the total number of patients in the country retained on NNRTI-based first-line regimens for 12 (or  $\geq 48$ ) months. The variance is calculated using linearization. A 95% confidence interval can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

## 10.5 Data analysis: Outcome 2a

Outcome 2a measures population-level retention at 12 months and the definition is consistent with that of the UNGASS/PEPFAR indicator. Data can be collected via a census of all patients in all sites. Alternatively, data can be collected via a survey within the same sites sampled for the measurement of viral suppression. If the latter is used, the clinic sampling weight is defined in Section 10.1. The patient sampling weight for clinic  $i$  is defined as the total number of eligible records

<sup>1</sup> STATA SURVEY DATA REFERENCE MANUAL RELEASE 13. 2013. StataCorp LP College Station, Texas, USA.

at that clinic,  $S_i$ , divided by the number of patient records reviewed at that clinic, excluding documented transfers-out. The overall weight is the product of the clinic and patient sampling weights. All patients, excluding documented transfers-out, are defined as either being retained (binary variable for retention = 1) or not retained (binary variable for retention = 0). The prevalence is estimated using a ratio. The numerator is an estimate of the total number of patients in the country who initiated treatment 12 months prior to the survey period who are retained on treatment. The denominator is an estimate of the total number of patients in the country who initiated treatment 12 months prior to the survey period. The variance is calculated using linearization. A 95% confidence interval can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

## 10.6 Data analysis: Outcome 2b

Outcome 2b measures viral load suppression (VL<1000 copies/mL) at 12 months among individuals sampled, adjusted for non-retention. This estimator assumes that all patients who are not retained on treatment at 12 months are not achieving viral load suppression.

$$\begin{aligned}\Pr(VLS) &= \Pr(VLS|Retained) \Pr(Retained) + \Pr(VLS|Not Retained) \Pr(Not Retained) \\ &= \Pr(VLS|Retained) \Pr(Retained) + 0\end{aligned}$$

The adjusted proportion of patients on treatment for 12 months with viral load suppression is estimated using a ratio.

The formulae are generalized for the setting with explicit stratification (strata indexed by  $h=1, \dots, H$ ), but they simplify readily when no stratification is used ( $H=1$ ). The clinic sampling weight is defined as in Section 10.1. The overall weight,  $w_{hi}$ , is the product of the clinic sampling weight and  $S_{hi}$  which is the number of eligible records for retention at clinic  $i$  in stratum  $h$ . The following formula for a ratio is used, where  $\hat{p}_{1,hi}$  is the clinic-specific estimate of the prevalence of viral load suppression at 12 months based on  $m_{hi}$  observations (patients with amplified VL), and  $\hat{p}_{RET,hi}$  is the clinic-specific estimate of the prevalence of retention at 12 months based on  $S_{hi}$  observations (records excluding documented transfers). The numerator is an estimate of the total number of patients in the country who initiated treatment 12 months prior to the survey period achieving viral load suppression. The denominator is an estimate of the total number of patients in the country who initiated treatment 12 months prior to the survey period.

$$\hat{V}_h = \sum_{i=1}^{n_h} w_{hi} \hat{p}_{1,hi} \hat{p}_{RET,hi}$$

$$\hat{V} = \sum_{h=1}^H \hat{V}_h$$

$$\hat{S}_h = \sum_{i=1}^{n_h} w_{hi}$$

$$\hat{S} = \sum_{h=1}^H \hat{S}_h$$

$$\hat{p}_2 = \left( \frac{\hat{V}}{\hat{S}} \right)$$

The formula for the linearized variance of this estimator is provided below, and it incorporates the finite population correction at three levels: (1) the clinic-level, (2) patient-level for the viral load suppression portion of the survey, and (3) record-level for the retention portion of the survey.

$$\begin{aligned}\widehat{var}(\widehat{V}) &= \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( w_{hi} \widehat{p}_{1,hi} \widehat{p}_{RET,hi} - \frac{\widehat{V}_h}{n_h} \right)^2 \\ &\quad + \sum_{h=1}^H \frac{n_h}{N_h} \sum_{i=1}^{n_h} w_{hi}^2 \left\{ \widehat{p}_{1,hi}^2 \left(1 - \frac{s_{hi}}{S_{hi}}\right) \frac{\widehat{p}_{RET,hi} (1 - \widehat{p}_{RET,hi})}{s_{hi}} \right. \\ &\quad + \widehat{p}_{RET,hi}^2 \left(1 - \frac{m_{hi}}{M_{hi}}\right) \frac{\widehat{p}_{1,hi} (1 - \widehat{p}_{1,hi})}{m_{hi}} \\ &\quad \left. - \left(1 - \frac{m_{hi}}{M_{hi}}\right) \frac{\widehat{p}_{1,hi} (1 - \widehat{p}_{1,hi})}{m_{hi}} \left(1 - \frac{s_{hi}}{S_{hi}}\right) \frac{\widehat{p}_{RET,hi} (1 - \widehat{p}_{RET,hi})}{s_{hi}} \right\}\end{aligned}$$

$$\widehat{var}(\widehat{S}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( w_{hi} - \frac{\widehat{S}_h}{n_h} \right)^2$$

$$\widehat{cov}(\widehat{V}, \widehat{S}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( w_{hi} \widehat{p}_{1,hi} \widehat{p}_{RET,hi} - \frac{\widehat{V}_h}{n_h} \right) \left( w_{hi} - \frac{\widehat{S}_h}{n_h} \right)$$

$$\widehat{var}(\widehat{p}_2) = \frac{1}{\widehat{S}^2} \{ \widehat{var}(\widehat{V}) - 2(\widehat{V}/\widehat{S}) \widehat{cov}(\widehat{V}, \widehat{S}) + (\widehat{V}^2/\widehat{S}^2) \widehat{var}(\widehat{S}) \}$$

This formula simplifies in the absence of stratification:

$$\begin{aligned}\widehat{var}(\widehat{p}_2) &= \left(1 - \frac{n}{N}\right) \frac{1}{\widehat{S}^2} \frac{n}{n-1} \sum_{i=1}^n w_i^2 (\widehat{p}_{1,i} \widehat{p}_{RET,i} - \widehat{p}_2)^2 \\ &\quad + \frac{1}{\widehat{S}^2} \frac{n}{N} \sum_{i=1}^n w_i^2 \left\{ \left(1 - \frac{s_i}{S_i}\right) \widehat{p}_{1,i}^2 \frac{\widehat{p}_{RET,i} (1 - \widehat{p}_{RET,i})}{s_i} + \left(1 - \frac{m_i}{M_i}\right) \widehat{p}_{RET,i}^2 \frac{\widehat{p}_{1,i} (1 - \widehat{p}_{1,i})}{m_i} \right. \\ &\quad \left. - \left(1 - \frac{s_i}{S_i}\right) \left(1 - \frac{m_i}{M_i}\right) \frac{\widehat{p}_{RET,i} (1 - \widehat{p}_{RET,i})}{s_i} \frac{\widehat{p}_{1,i} (1 - \widehat{p}_{1,i})}{m_i} \right\}\end{aligned}$$

A 95% confidence interval can be calculated in the following way:  $\widehat{p}_2 \pm t_{n-H, 0.975} \sqrt{\widehat{var}(\widehat{p}_2)}$

The above interval is a Wald-type interval for the proportion. Other options for confidence interval estimation include transforming the interval to a logit scale or calculating a Wilson interval using the effective sample size.

## 10.7 Data analysis: Outcome 3a

Outcome 3a measures the prevalence of HIVDR among individuals sampled on ART for 12±3 (or ≥48) months with VL ≥ 1000 copies/ml. Outcome 3a is a subpopulation analysis of the overall data because the population is restricted to those individuals without viral load suppression.

The clinic sampling weight is defined in Section 10.1. The patient sampling weight is defined in Section 10.2. For all HIVDR outcomes, we must also define a non-response weight to compensate for genotyping failure. For all individuals with observed genotype, their non-response sampling weight is defined as the number of patients with observed VL failure at their clinic divided by the number of patients with observed VL failure and observed genotype at their clinic. The non-response weight assumes that genotyping failure is unrelated to the presence of HIVDR mutations. For all individuals with missing genotype, their non-response sampling weight is missing. For all individuals with viral load suppression, their non-response weight is equal to 1. The overall weight is the product of the clinic, patient and non-response sampling weights.

The population is restricted to patients without viral load suppression using the subpopulation command in Stata. All patients without viral load suppression with available genotypes are defined as either having HIVDR (binary variable for any HIV drug resistance mutation = 1) or no documented HIVDR (binary variable for any HIV drug resistance mutation = 0). The prevalence is estimated using a ratio. The numerator is an estimate of the total number of patients on ART for 12±3 (or ≥48) months failing to achieve viral load suppression and with detected HIVDR mutations. The denominator is an estimate of the total number of patients on ART for 12±3 (or ≥48) months in the country retained on treatment for 12±3 (or ≥48) months failing to achieve viral load suppression. The variance is calculated using linearization. A 95% confidence interval can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

## 10.8 Data analysis: Outcome 3b

Outcome 3b measures the prevalence of HIVDR among individuals sampled on first-line regimens with VL ≥ 1000 copies/ml receiving ART for 12±3 (or ≥48) months. Outcome 3b is a subpopulation analysis of the data used for Outcome 3a. Data analysis is conducted using the same sampling weights described for Outcome 3a. The population is restricted to patients on first-line regimens using the subpopulation command in Stata, in addition to the restrictions described for Outcome 3a. All patients on first-line regimens without viral load suppression with available genotypes are defined as either having HIVDR (binary variable for any HIV drug resistance mutation = 1) or no documented HIVDR (binary variable for any HIV drug resistance mutation = 0). The prevalence is estimated using a ratio. The numerator is an estimate of the total number of patients on ART for 12±3 (or ≥48) months receiving first-line regimens, failing to achieve viral load suppression and with detected HIVDR mutations. The denominator is an estimate of the total number of patients on ART for 12±3 (or ≥48) months receiving first-line regimens and failing to achieve viral load suppression. The variance is calculated using linearization. A 95% confidence interval can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

## 10.9 Data analysis: Outcome 3c

Outcome 3c measures the prevalence of HIV drug resistance among individuals sampled on NNRTI-based first-line regimens with viral loads ≥ 1000 copies/ml receiving ART for 12±3 (or ≥48) months. Outcome 3c is a subpopulation analysis of the data used for Outcome 3a. Data analysis is conducted using the same sampling weights described for Outcome 3a. The population is restricted to patients on NNRTI-based first-line regimens using the subpopulation command in Stata, in addition to the restrictions described for Outcome 3a. All patients on first-line regimens without viral load suppression with available genotypes are defined as either having HIVDR (binary variable for any HIV drug resistance mutation = 1) or no documented HIVDR (binary variable for any HIV drug resistance mutation = 0). The prevalence is estimated using a ratio. The numerator is an estimate of the total number of patients on ART for 12±3 (or ≥48) months receiving NNRTI-based first-line regimens, failing to achieve viral load suppression and with detected HIV drug resistance mutations. The denominator is an estimate of the total number of patients on ART for 12±3 (or ≥48) months receiving NNRTI-based first-line regimens and failing to achieve viral load suppression. The variance is calculated using linearization. A 95% confidence interval can be calculated using a standard Wald formula or by a logit transformation (default in Stata).

## 10.10 Data analysis: Outcome 4

Outcome 4 is the prevalence of HIVDR among all individuals sampled on ART for 12±3 (or ≥48) months. Data analysis is conducted using the same sampling weights described for Outcome 3a, though the population is not restricted for Outcome 4. All patients with observed VL measurements are defined as either having HIVDR (binary variable for any HIV drug resistance mutation = 1) or no documented HIVDR (binary variable for any HIV drug resistance mutation = 0). All patients with viral load suppression are assumed to have no documented HIVDR. The prevalence is estimated using a ratio. The numerator is an estimate of the total number of patients on ART for 12±3 (or ≥48) months retained on treatment in the country failing to achieve viral load suppression and with detected HIVDR mutations. The denominator is an estimate of the total number of patients on ART for 12±3 (or ≥48) months retained on treatment in the country. The variance is calculated using linearization. A 95% confidence interval can be calculated using a standard Wald formula or by a logit transformation (default in Stata).





# Notes

Lined writing area with horizontal dotted lines.





**For more information, contact:**

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

E-mail: [hiv-aids@who.int](mailto:hiv-aids@who.int)

[www.who.int/hiv](http://www.who.int/hiv)

ISBN 978 92 4 150707 3



9 789241 507073