UNAIDS/WHO Working Group
on Global HIV/AIDS and STI surveillance

Guidelines for measuring national HIV prevalence in population-based surveys

World Health Organization

Joint United Nations Programme on HIV/AIDS
UNAIDS
UNHCR • UNICEF • WFP • UNDP • UNFPA
UNODC • ILO • UNESCO • WHO • WORLD BANK

Guidelines for measuring national HIV prevalence in population-based surveys

UNAIDS/WHO Working Group
on Global HIV/AIDS and STI surveillance

Joint United Nations Programme on HIV/AIDS
UNAIDS
UNHCR • UNICEF • WFP • UNDP • UNFPA
UNODC • ILO • UNESCO • WHO • WORLD BANK

World Health Organization

CONTENTS

ACKNOWLEDGEMENTS

Global surveillance of HIV/AIDS and sexually transmitted infections is a joint effort of the World Health Organization (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS). The UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, initiated in November 1996, is the main coordination and implementation mechanism for UNAIDS and WHO to compile the best information available and to improve the quality of data needed for informed decision-making and planning at the national, regional and global levels.

WHO and UNAIDS would like to thank Godwill Asiimwe-Okiror, Rodney Knight and Eleanor Gouws for preparing this document.

Significant contributions were also received from G.M. Gershy-Damet and Emil Asamoah-Odei, WHO Regional Office for Africa; Jesus M. Carcia Calleja and George Loth, WHO headquarters; Peter Ghys, Rob Lyerla and Karen Stanecki, UNAIDS headquarters; Bernard Barrere, Ann Way, Annie Cross and Alfredo Fort, ORC Macro International; Tim Dondero, Theresa Diaz and Stefan Wiktor, United States Centers for Disease Control and Prevention; John Novak, United States Agency for International Development; Basia Zaba, London School of Hygiene and Tropical Medicine; John Stover, Futures Group International; Gail Henderson, University of North Carolina; Thomas Rehle, independent consultant in international health and disease control, Washington, DC; Anne Buvé, Institute of Tropical Medicine, Antwerp, Belgium; and Olive Shisana, Human Sciences Research Council, South Africa.

The concepts and material presented in this publication were informed by a workshop organized by UNAIDS and WHO in Lusaka, Zambia in 2002 and by meetings of the UNAIDS Reference Group on Estimates, Modelling and Projections in 2003 and 2004. WHO and UNAIDS would like to acknowledge the contributions of the participants of these meetings, including staff of national AIDS programmes and experts in various fields related to HIV epidemiology.

MEASURE DHS+ is a project funded by the United States Agency for International Development and implemented by ORC Macro International, Calverton, MD, USA, and had contributed globally to the implementation of national household surveys.

ABBREVIATIONS

AIDS	acquired immunodeficiency syndrome
DHS	demographic and health survey
ELISA	enzyme-linked immunosorbent assay
EPP	Estimation and Projection Package
HIV	human immunodeficiency virus
MEASURE	Monitoring and Evaluation to Assess and Use Results
NHANES	National Health and Nutrition Examinations Surveys
STI	Sexually transmitted infections
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNICEF	United Nations Children's Fund
WHO	World Health Organization

1. INTRODUCTION

Purpose

The purpose of this publication is to provide guidelines for countries planning to measure the national prevalence of HIV infection through population-based surveys. Guidelines are provided on three main topics:

- conducting a new national population-based survey to measure HIV prevalence;
- incorporating HIV testing into existing national household surveys or demographic and health surveys; and
- calculating a national estimate of HIV prevalence based on data obtained from population-based surveys and sentinel surveillance.

The guidelines have been developed to assist surveillance officers and programme managers involved in HIV/AIDS surveillance activities in planning and conducting population-based HIV prevalence surveys. The document describes the various processes involved in planning a survey (including survey design, sampling methods, HIV strategies, questionnaire design, budget planning, ethical considerations and survey operations) and conducting a population-based survey (including field operations, HIV testing and data management). Finally, it provides guidelines on how to analyse and reconcile the results obtained from national population-based surveys with those obtained from sentinel surveillance to produce an estimate of HIV prevalence in a country.

The purpose of this document is not to provide detailed instructions on how to conduct population-based surveys but rather to give an overview of the major issues that need to be considered when planning and conducting such surveys. More detailed instructions on how to implement population-based surveys are available from various sources, such as the UNICEF (2000) manual on multiple indicator cluster surveys or the model manuals on how to implement demographic and health surveys (MEASURE DHS+, 2005a), which can also be used to train local staff.

Background

Obtaining an estimate of the number of people infected with HIV in a country or region is important for the purpose of evaluation, programme planning and advocacy. Estimates are usually obtained from surveillance systems, but because HIV epidemics develop differently in different countries, different surveillance systems are required (UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, 2000). In all epidemic states, however, surveillance systems aim to provide information that will increase and improve the response to the HIV epidemic. In countries where HIV is uncommon, biomedical surveillance and behavioural data can provide an early warning of a possible epidemic. Where HIV is concentrated in subgroups with high-risk behaviour, surveillance can provide valuable information for designing focused interventions. In generalized epidemics, sentinel HIV surveillance among the general population can provide essential information for planning care and support and for indicating the success of the current response (UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, 2000).

In the mid-1980s, WHO recommended facility-based sentinel surveillance of HIV for monitoring the HIV epidemic, mainly because of easy access to people attending public health facilities (Chin & Mann, 1989). In most countries with generalized epidemics, annual HIV surveillance among pregnant women attending public-sector antenatal clinics has since become

the primary source of data on the spread of HIV (UNAIDS/WHO, 2003b), supplemented in some cases by surveys among people attending sexually transmitted infection clinics. The prevalence among pregnant women approximates well the prevalence among sexually active men and women aged 15–49 years (Grassly et al., 2004; UNAIDS/WHO, 2003b). As the epidemic has spread and more resources have become available, countries have expanded the surveillance system by increasing the number of sites in urban as well as rural areas.

Concerns about the representativeness and accuracy of national HIV estimates derived from antenatal clinic surveillance have led to an increased demand for more surveys and more data on the prevalence and distribution of HIV in the whole population. In recent years, several countries have included HIV testing in national population-based surveys. Technological developments, such as the use of dried blood spots for collecting HIV samples and rapid HIV testing, have greatly facilitated the collection of biological data in population-based surveys.

Large-scale population-based surveys and sentinel surveillance both have weaknesses, but combining the two sources of data will yield more accurate estimates of HIV prevalence. These guidelines show how to obtain a national estimate of HIV by adjusting prevalence measures from population-based surveys to reduce bias when needed and by adjusting prevalence measures from sentinel surveillance systems using the results obtained from the population-based survey (either adjusted or unadjusted).

Antenatal clinic surveillance

The main purpose of surveillance based on women attending antenatal clinics is to assess trends in HIV prevalence over time. However, because other data sources are lacking, antenatal clinic surveillance has also been used to estimate the population levels of HIV.

HIV surveillance in antenatal clinics has been implemented in more than 115 countries worldwide (UNAIDS/WHO, 2003b). This is usually based on anonymous, unlinked, cross-sectional surveys of pregnant women attending antenatal clinics in the public health sector. Only first-time attendees are included to minimize the chance of any woman being included more than once. Blood is taken routinely from pregnant women for diagnostic purposes (such as to test for syphilis or rhesus factor). After personal identifiers are removed (except for a few key characteristics such as age, parity, marital status and clinic location (or urban versus rural residence)), the blood is tested for HIV. Antenatal clinic surveys are usually done annually at the same time of the year to obtain an estimate of the point prevalence for that year but also to assess trends over time. More detail can be found in the UNAIDS/WHO (2003b) guidelines for conducting HIV sentinel surveys among pregnant women.

Strengths

- Antenatal clinics provide ready and easy access to a cross-section of sexually active women from the general population who are not using contraception. Blood is drawn for routine testing for syphilis, and a portion can be used for anonymous testing of HIV.
- In generalized epidemics, HIV testing among pregnant women is considered a good proxy for prevalence in the general population.
- Annual antenatal clinic survey data can be used to assess trends in the HIV epidemic over time.
- Data for pregnant women will reflect the prevalence in groups that may be of higher risk of infection because of their living arrangements (such as workers who live in

hostels or army barracks) if they have regular unprotected sexual contact with women in the general population.

- The limitations of antenatal surveillance are recognized and acknowledged, and where possible, correction factors have been developed to overcome some of the limitations.
- In countries with low levels of HIV prevalence, strategically placed sentinel sites can provide an early warning for the start of an epidemic.
- In recent years, many countries have expanded the geographical coverage (the number and sample sizes of sites) of sentinel surveillance, especially in rural areas, to improve the representativeness of the samples.

Weaknesses

- Most sentinel surveillance systems have limited geographical coverage, especially in smaller and more remote rural areas.
- Women attending antenatal clinics may not be representative of all pregnant women because many women may not attend antenatal clinics or may attend private clinics.
- The rate of contraceptive use in a country may affect the number of pregnant women.
- The implementation of antenatal clinic–based surveillance varies considerably between countries (Garcia-Calleja et al., 2004). The quality of the surveys may vary over time depending on available resources.
- Antenatal clinic surveillance does not provide information about HIV prevalence in men. Because these surveys are conducted among pregnant women, estimates for men are based on assumptions about the ratio of male-to-female prevalence that are derived from community-based studies in the region. However, this ratio varies between countries and over time.

Population-based surveys including HIV testing

The demand by decision-makers for better data on the burden of HIV/AIDS in countries and the limitations of antenatal surveillance systems with respect to geographical coverage, under-representation of rural areas and the absence of data for men have led to an interest in including HIV testing in national population-based surveys. Population-based surveys can provide reasonable estimates of HIV prevalence for generalized epidemics, where HIV has spread throughout the general population in a country. However, for low-level and concentrated epidemics, these surveys will underestimate HIV prevalence, because HIV is concentrated in groups with high-risk behaviour and these groups are usually not adequately sampled in household-based surveys. For example, the HIV prevalence obtained from a national household-based survey in the United States [National Health and Nutrition Examination Survey (NHANES)] conducted between 1999 and 2001 underestimated HIV prevalence by about 30%. By the nature of its design, NHANES is limited to households and therefore excludes some groups (such as military personnel, prison populations and homeless people). The NHANES estimate of HIV prevalence was 0.43% among people aged 18–49 years (McQuillan et al., 2005) versus the United States Centers for Disease Control and Prevention and UNAIDS estimate for the same period of about 0.6%, based on an analysis of the HIV and AIDS case surveillance systems of the United States Centers for Disease Control and Prevention (1999) to estimate the number of people living with HIV and AIDS (UNAIDS, 2004a).

In recent years, the number of population-based surveys that collect biological specimens for HIV testing has increased. Many of these surveys covered women and men of

reproductive ages (women 15–49 years old and men 15–54 years old) and used dried blood spots for collecting specimens (Table 1 presents the reported surveys). Some early surveys were designed for unlinked anonymous testing, in which the HIV test results could not be linked to individuals, whereas more recent surveys have incorporated linked anonymous testing, in which HIV test results can be linked to behavioural data without revealing the identity of any individual who has been tested.

Strengths

- In generalized epidemics, population-based surveys can provide representative estimates of HIV prevalence for the general population as well as for different subgroups, such as urban and rural areas, women and men, age groups and region or province.
- The results from population-based surveys can be used to adjust the estimates obtained from sentinel surveillance systems.
- Population-based surveys provide an opportunity to link HIV status with social, behavioural and other biomedical information, thus enabling researchers to analyse the dynamics of the epidemic in more detail. Information from this analysis could lead to better program design and planning.

Weaknesses

- In population-based surveys, sampling from households may not adequately represent high-risk and mobile populations. In low-level or concentrated epidemics, population-based surveys therefore underestimate HIV prevalence.
- Nonresponse (either through refusal to participate or absence from the household at the time of the survey) can bias population-based estimates of HIV. (Collecting information on nonresponders can help in the process of adjusting for nonresponse.)
- Population-based surveys are expensive and logistically difficult to carry out and cannot be conducted frequently. Typically, these surveys are conducted every 5–10 years.

Combining data from sentinel and population-based surveys

Sentinel surveillance and population-based surveys each have strengths and weaknesses but together provide complementary information. Sentinel surveillance provides samples that are consistent over time so that good estimates of HIV trends can be obtained. They can also provide good overall national coverage and allow estimates to be generated by age and geographical location. Population-based surveys, in contrast, provide much better coverage of the general population, including men, and can provide much more detailed information on social, economic and sexual behaviour and biomedical factors associated with HIV infection. Because of the cost, they can usually not be conducted regularly and therefore provide limited temporal coverage. However, taken together, sentinel surveillance and population-based surveys can provide a clear picture of both overall trends and geographical distribution of HIV as well as detailed information on potential risk factors and groups exhibiting high-risk behaviour.

Guidelines for conducting antenatal clinic surveys are available (UNAIDS/WHO, 2003b) and are not discussed in this publication. Here we focus on incorporating HIV testing into national population-based surveys (when planning a new study and when incorporating HIV testing into existing surveys) and how to obtain national estimates of HIV prevalence from data obtained from both sentinel surveys and population-based surveys.

Table 1. Population-based HIV prevalence surveys (with results available as of February 2005)

Country	Survey year	Survey type	Study population (age in years for women and men)	Specimen collected	Linking of HIV test to survey data
Burkina Faso	2003	Demographic and health survey	15–49 women 15–59 men	Dried blood spot	Linked anonymous
Burundi	2002	HIV/AIDS	12 years and above	Venous blood	Linked anonymous
Cameroon	2004	Demographic and health survey	15–49 women 15–59 men	Dried blood spot	Linked anonymous
Congoa	2003	HIV/AIDS	15–49, women and men	Venous blood	Linked anonymous
Dominican Republic	2002	Demographic and health survey	15–49 women 15–59 men	Saliva (oral transudate)	Unlinked anonymous
Ghana	2003	Demographic and health survey	15–49 women 15–59 men	Dried blood spot	Linked anonymous
Kenya	2003	Demographic and health survey	15–49 women 15–59 men	Dried blood spot	Linked anonymous
Mali	2001	Demographic and health survey	15–49 women 15–59 men	Dried blood spot	Unlinked anonymous
Niger	2002	HIV/AIDS	15–49, women and men	Dried blood spot	Unlinked anonymous
Rwanda	1997–1998	HIV/AIDS	12–49, women and men	Dried blood spot	Linked anonymous
Sierra Leone	2002	HIV/AIDS	12–49, women and men	Dried blood spot	Unlinked anonymous
South Africa	2002	HIV/AIDS	2 years and above	Saliva (oral transudate)	Linked anonymous
South Africa	2003	HIV/AIDS	15–24 years	Saliva (oral transudate)	Linked anonymous
Tanzania	2003	AIDS indicator survey	15–49, women and men	Dried blood spot	Linked anonymous
			0–4 years	Dried blood spot	
Zambia	2001–2002	Demographic and health survey	15–49 women 15–59 men	Dried blood spot	Unlinked anonymous
Zanzibar	2002	HIV/AIDS	10 years and above	Dried blood spot	
Zimbabwe	2001–2002	Young adults, HIV/AIDS	15–29, women and men	Dried blood spot	Linked anonymous

[a]Only the urban population.

2. Planning a population-based survey to measure HIV prevalence

Countries that do not have existing population-based surveys to which HIV testing can be added have to design and conduct new seroprevalence surveys. Designing and successfully implementing a population-based survey to measure HIV prevalence requires that the agency implementing the survey have considerable experience in conducting large-scale national surveys.

Careful planning is an important phase in the survey process because everything must be in place at the right time for successful implementation. The duration of this phase will vary from country to country but could take up to six months or more. Organizational capacity and commitment and availability of funds all influence the duration of the planning phase.

The planning phase usually starts with the organizing body writing a concept paper about the population survey, indicating the rationale for conducting the survey and the assistance required to conduct the survey. The concept paper should be discussed through consultative meetings with key stakeholders and development partners, to solicit their technical and financial support. It should also be discussed with key ministries and offices such as the statistics office and ministries responsible for health, finance and local government. The finance ministry may be needed to provide funding or waive taxes on items required for the survey. The survey will use the administrative structures from the national level to the grassroots level during the mobilization stage, and getting support from the relevant government bodies during the early planning stages is essential.

Once the concept paper has been approved, a steering committee and technical working group will be selected and will develop the survey protocol. Technical assistance in developing the protocol may be sought from local universities or research organizations or from organizations such as WHO, the United States Centers for Disease Control and Prevention, UNAIDS or MEASURE DHS+ (a project funded by the United States Agency for International Development and implemented by ORC Macro International, Calverton, MD, USA). The objectives, design and survey methods have to be clearly outlined in the proposal. Part of the protocol development process is to create a budget for the survey, because the cost of the survey often affects the design. The survey protocol should also specify the laboratories that will carry out the HIV testing, including external quality control. Identifying these laboratories might require visiting selected laboratories to assess their capacity and requirements. The protocol should then be submitted to national (and, where needed, international) ethics boards or committees for review and approval.

After the protocol is approved but before the survey begins, the necessary equipment and supplies (laboratory equipment, reagents, other laboratory supplies, transport, communication equipment, computer equipment and other survey supplies) have to be procured. Formal arrangements might have to be made to hire vehicles and communication equipment, and the data management unit should also be set up during this phase. Any memorandum of understanding between the coordinating organization and the various agencies and organizations providing financial and or technical assistance should be written and signed during this phase. The memorandum of understanding should identify the responsibilities and commitment of each of the organizations involved. The coordinating organization might have to work with the central statistics office or population census office to obtain the sampling frames and maps necessary for planning the sampling methods and selecting the actual samples.

Informed consent forms, questionnaires and supervision forms are developed during the planning stage. Questionnaires have to be translated into the local languages, and

translations have to be verified at this stage. The questionnaire and procedures for collecting, storing and testing specimens are pretested during the planning stage to verify the study design and procedures.

Steps to consider when planning a population-based survey are outlined in Box 1 and discussed below.

Box 1. Steps involved in planning a population-based HIV seroprevalence survey

1. Define the objectives of the survey
2. Specify the design of the survey
3. Decide on the geographical area, survey population and timing of the survey
4. Design sampling methods and clearly define:
   - the target population
   - the strata and/or clusters to include in the sample (region, urban or rural and the enumerating area)
   - sampling units (such as households or individuals)
5. Calculate the sample size
6. Decide on the data to be included in the survey:
   - Demographic data
   - sexual behavioural data, knowledge related to HIV and reproductive health
   - other biomarkers
7. Design the questionnaire for data collection
8. Decide on the HIV testing strategy:
   - blood, saliva or urine
   - type of testing and specific assays
   - procedures for confirmatory testing
   - quality assurance for the laboratory
9. Plan the survey budget
10. Uphold ethical principles and obtain ethical approval for the survey
11. Design informed consent procedures
12. Plan the survey operations
    - staff organization
    - mobilization
13. Prepare manuals and pretest questionnaires and HIV testing procedures
14. Train the staff

2.1. Survey objectives

The objectives of the survey have to be clearly identified during the planning phase, including the timing of the survey, the geographical area and the age of the participants. Both the broad objectives and country-specific priorities should be described.

The overall, primary objective of a population-based HIV prevalence survey is typically to obtain accurate HIV prevalence estimates and information on risk factors related to HIV infection at the national and subnational levels that will inform the design, implementation and evaluation of the national response to the HIV/AIDS epidemic.

The secondary objectives of conducting a population-based HIV prevalence survey could include:

- to obtain national and subnational estimates of HIV prevalence and demographic variation in HIV prevalence in the general population;
- to provide information on HIV prevalence that can be used to calibrate, validate and improve the use of HIV sentinel surveillance data among pregnant women attending antenatal clinics;
- to identify risk factors that predispose the general population and subpopulations to HIV infection;
- to link the risk factors with biological measures and to assess the associations between the two; and
- to assess the extent to which current interventions are accessible to the general population, including their impact on the population.

2.2. Survey design

The design of the study is important for planning the survey and sampling methods. Population-based HIV prevalence surveys usually follow a cross-sectional design with the general population as the study subjects. A cross-sectional study usually provides a "snapshot" of the frequency and characteristics of a disease in a population at a specific time.

Most often, surveys are designed to estimate a key indicator such as HIV prevalence, at one point in time: to obtain a measure of the current level of the epidemic. However, an objective of the survey could also be to measure change in HIV prevalence over time, which will affect the design and methods of the study.

2.3. Geographical area, survey population and timing

A decision must be made about whether to carry out a national survey (in which all eligible people in the country are included) or to conduct the survey in a restricted area. This decision usually depends on the size of the area and the available resources. In a small country, all districts may be sampled. In a large country, however, this might not be possible, but a random sample of districts can be selected for inclusion in the study.

In addition to specifying the geographical area to be included in the survey, the features of the study population have to be clearly specified, such as whether the survey is restricted to a particular sex, cultural or age group. To obtain estimates of HIV in a country where the general mode of transmission is heterosexual, the study population should ideally include both males and females. The demographic and health household surveys typically collect data on women 15–49 years old and on men 15–54 years old.

Exclusion criteria generally include people in special institutions such as hospitals, prisons, schools, hotels, military camps and refugee camps, or homeless people, because of the complexities associated with enrolling such people in the survey.

Inclusion of children aged 0–14 years

Countries conducting national surveys to obtain measures of HIV prevalence have to decide whether or not to include children in the survey, and such a decision could be guided by the local epidemiology of HIV. Available epidemiological data in most countries indicate low HIV prevalence among children aged 5–14 years, although high prevalence rates have been reported from South Africa (Shisana et al., 2002). Because the prevalence of HIV in this age

group is low in most countries, including children 5–14 years old in population-based surveys may not be cost-effective.

Including children younger than 18 months requires using technologies that identify HIV antigens or HIV viral nucleic acid or the virus itself to confirm the HIV status of the child. These tests are expensive, technically more difficult to perform and more often likely to be affected by laboratory error (UNAIDS/WHO Working Group on Surveillance, 2001). Many countries currently lack the infrastructure to carry out these tests.

Timing of the survey

Issues to consider for timing of the survey include the following.

- Repeat surveys (to assess change in HIV prevalence over time): surveys should be conducted at regular, fixed intervals (such as every five years) to provide insight into trends.
- Local holidays and festivals: surveys should not usually be scheduled to take place during local holidays or festivals when people often travel or are likely to be away from their homes. However, many migrant workers return to their homes over holiday periods and more men may be available to participate during these periods.
- Time of day: if the aim is to include both men and women in the survey, the best time to do household visits may be in the evening (or over weekends) when people are back from work or school. The danger of conducting surveys during official working hours is that many people, especially men, are away from home. If the survey is conducted outside the usual working hours, it is advisable to inform people in advance of the date and time the household will be visited (and, if possible, to make a prior appointment with the family).
- Plan to allow for additional survey weeks in case extra time is needed.

2.4. Sampling methods

Basic concepts of sampling

Studying everybody in a population of interest is often not practical or feasible. Selecting a sample that is representative of the population can provide reliable information from which inferences can be made about the entire population. The main aim when selecting a sample is therefore to ensure that the sample closely resembles the population of interest. Obtaining a representative sample usually requires selecting a probability sample: one in which each element in the population has a known and non-zero probability of being included in the sample.

The four types of probability sample most often used in population surveys include simple random, systematic, stratified and cluster sampling (Stroup & Teutsch, 1998).

In simple random sampling, a sample of size n is randomly selected from a population of size N and every element in the population has an equal probability (n/N) of being selected.

In systematic sampling, the first element is selected randomly and thereafter every kth element is selected, with $k = N/n$. Systematic and simple random sampling require a list of the population to be used as the sampling frame from which the samples will be selected.

Stratified sampling is accomplished when the population of interest is divided into appropriate non-overlapping sub-populations or strata (such as province or geographical area

(urban or rural)) and a probability sample is selected from within each stratum. Administrative geographical areas, such as regions or districts, can form natural divisions for strata. Within geographical areas, strata can often be further divided into urban and rural areas.

Cluster sampling is the method most often used in public health surveys when the travel and personnel costs associated with conducting surveys are important factors. It is considered the most feasible and economic sampling method in large surveys (Hulley & Cummings, 1988; Lemeshow et al., 1990). In a cluster design, the population is divided into groups (or clusters) of elements, called sampling units (such as enumeration areas, villages or city blocks or households). A sample of clusters is then selected randomly from the population, and either all the elements or a sample of elements is selected from each of the clusters. An advantage of cluster sampling is that, once the clusters are selected, only a list of all the elements from those selected clusters is needed, which can reduce time and expense.

Multi-stage sampling is generally used in population-based surveys and incorporates several of the above sampling strategies together. For example, to assess the adult prevalence of HIV at a national level, the first step is to divide the population into strata such as provinces or urban and rural districts, from which clusters such as villages or city blocks are randomly selected, and from which households can finally be randomly selected. From these households, all individuals 15–49 years old can then be selected for inclusion in the study. For national population-based surveys, the most recent population census in the country often serves as a sampling frame from which to select the sampling units. Most census systems also have enumeration areas that can be grouped into geographical and urban or rural areas. If a census is not available, another form of population information, such as voter registration, can be used to select the sample.

The probabilities of selecting a sample must be noted at every stage of the design, and weights must be applied when the probabilities of selection differ for the elements in the survey to avoid introducing bias. A sampling weight is defined as the inverse of the probability of selection. The sum of the sampling weights of the selected elements must equal the total population size (Stroup & Teutsch, 1998). When data from a complex sampling survey are analysed, the design of the survey must be taken into account in the analysis (including weights when required).

Probability proportional to size sampling, in which the probability of selecting a particular sampling unit will be proportional to its population size, is commonly used for selecting the clusters in multi-stage cluster sampling. Probability proportional to size is self-weighting and simplifies the analysis of data (weighting is not required in the analysis).

The design effect (DEFF) should be considered when a complex sample survey is proposed. The design effect is defined as the ratio of the variance when a complex sample design is used to the variance that would be expected for a simple random sample of the same size. Considering the design effect in a survey is important because it provides a measure of efficiency of the design. By definition, the design effect is 1 for simple random sampling. A design effect greater than 1 (which is expected for most complex designs) means that the sampling design reduces the precision[1] of estimates compared with simple random sampling: that is, samples with design effects >1 will yield variances that are higher than comparable simple random samples. Consequently, when a study is designed using a complex design, the sample size has to multiplied by the design effect to obtain the desired level of precision.

The design effect can be calculated as

$$\text{DEFF} = 1 + (m - 1)\rho$$

[1] The precision of survey results refers to how closely the results from a sample can reproduce the results that would be obtained from a complete count (census) conducted using the same techniques.

where m is the number of elements selected in each cluster and ρ is the intraclass (or intracluster) correlation coefficient, defined as $\rho = \sigma_c^2/(\sigma^2 + \sigma_c^2)$ (or the ratio of between-cluster variability to total variability) (Som, 1973).

When a survey is conducted using cluster sampling, the factor $(m - 1)\rho$ gives a measure of the relative change in the sampling variance due to sampling clusters instead of sampling the elementary units directly.

For example, if clusters of size $m = 10$ are sampled and $\rho = 0.1$, then $1 + (m - 1)\rho = 1.9$, so that the variance of cluster sampling is about twice that of simple random sampling of individuals. Thus, even a small intracluster correlation coefficient could lead to a substantial increase in variance.

Issues to consider when planning sampling methods

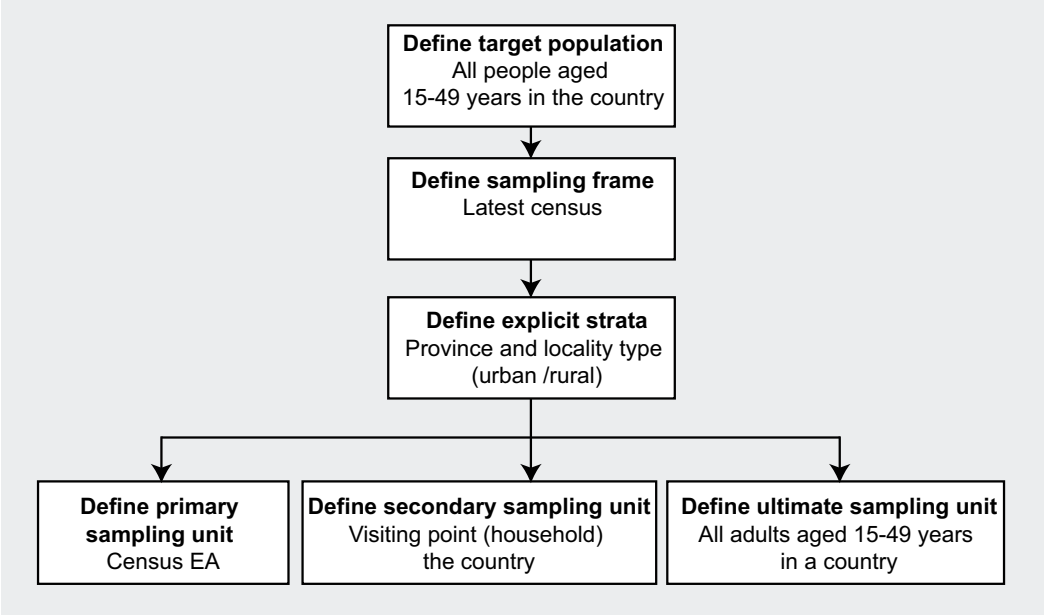
Planning and conducting a population-based survey should be a collaborative effort, and the lead implementing body should include statisticians as part of the survey personnel, either from within the country (such as from the central bureau of statistics or from universities or research institutions) or, if necessary, from outside the country. Statisticians should ideally have experience in survey design for population-based surveys to assist with sampling methods and sample size calculations.

The types of sampling methods used in a survey depend on the objectives and scope of the survey, the overall survey budget, the method of data collection and the participants needed. A first step, however, is to clearly define the relevant target population. The target population (often referred to as the "population") refers to the total population for which the information of interest (in this instance HIV status) is required. The target population can be all the people in the entire country or all the people in a certain province or region, or it can be a subset such as all adults 15–49 years old in a given location. The definition of the target population therefore should include the geographical area (country or region) as well as the age group (and sex) of interest.

Basic information about the sampling frame (the tool used to gain access to the population) should be provided. The sampling frame can consist of a list of geographical areas (area frame) that provide indirect access to individuals (such as the neighbourhoods in a city). However, the most commonly used sampling frame for a country is the national population registry or the census enumeration areas from the census bureau. In multistage surveys, a different sampling frame is used at each stage of the sampling. The units listed in the frame are generally called the sampling units. In large population surveys, primary (enumeration area), secondary (visiting point, such as household or apartment block) and ultimate sampling units (individuals) are identified. The sampling units for the first stage of a multistage design are called primary sampling units, and the sampling units for the final stage are called enumeration areas or ultimate units.

To give an example of a multistage sampling design: enumeration areas, or the primary sampling units, can be selected from each geographical stratum using probability proportional to size of the enumeration area. The secondary sampling units, households, can then be selected from each of the selected enumeration areas using simple random sampling or systematic sampling from a full list of the households of the selected enumeration areas. Finally, the predetermined number of ultimate sampling units (eligible respondents) can be selected from a full list of all eligible respondents in each of the selected households using simple random sampling. Fig. 1 shows a schematic diagram of various steps in the sample design.

Figure 1: Example – steps in the sample design



2.5. Calculating sample size

When surveys are designed, one of the most important considerations is to ensure that the estimates obtained from the survey will be reliable[2] enough to meet the objectives of the survey (Lemeshow et al., 1990). Reliability is associated with sample size and, in general, a larger sample will result in greater reliability. Specifying the number of sampling domains[3] and specifying the desired level of reliability for estimates at the domain level are therefore important first steps in estimating the sample size.

In general, the primary objective of a survey is to estimate an unknown population parameter with a given precision. Here the aim is to estimate the prevalence (p) of HIV with a given precision. When a sample is taken, irrespective of the type of sample, or whether the sampling takes place in one or multiple stages, an estimate of the prevalence ($\hat{p}$) is produced together with an estimate of the variance (Var ($\hat{p}$)). One way of expressing the desired precision of the estimate is to specify that the estimated prevalence should not differ from the real prevalence by more than d percentage points with 95% confidence.

For *simple random sampling*, the population prevalence can be estimated by

$$\hat{p} = \sum_i \frac{y_i}{n}$$

where $\sum y_i$ is the total number of all cases testing positive for HIV and n is the total number of people who were tested, and

$$Var\,(\hat{p}) = \frac{p(1-p)}{n}$$

[2] One essential element of the sampling plan and estimation procedure is that it should provide estimates of HIV for which the mean of the sampling distribution is equal to or at least close to the true unknown population prevalence and for which the standard error is small.

[3] A sampling domain is the geographical area for which an estimate is desired.

To estimate p within "d" percentage points, the sample size can be calculated as

$$n = \frac{z^2 p(1-p)}{d^2}$$

where z is the critical value of the confidence interval for a standard normal distribution (for 95% confidence intervals, $z = 1.96$), p is an estimate of the expected prevalence and d depends on the required precision as noted above).

Example: a simple random sample is desired to estimate the HIV prevalence in a town, such that the estimated prevalence will be within 2 percentage points of the true value with 95% confidence. The sample size can be calculated as follows (assuming that the prevalence in the village is about 10%):

$$n = \frac{(1.96)^2 0.1(1-0.1)}{0.02^2} = 864.4,$$

suggesting that a total sample of 865 should be studied.

For *stratified sampling* the population is divided into L strata, and random samples have to be selected from each stratum.

The proportion of individuals in the population who are infected with HIV is calculated as

$$p = \sum_{i=1}^{L} p_i \frac{N_i}{N}$$

where p_i is the proportion of individuals in stratum i that are infected with HIV, N_i is the population size in stratum i and N is the total population size: that is, the population proportion is a weighted average of the stratum-specific proportions, where the weights are the relative sizes of the strata.

To estimate the sample size (for stratified sampling) required to estimate p within d percentage points with 95% confidence, the following formula can be used (Lemeshow et al., 1990).

$$n = \frac{z^2 \sum_{i=1}^{L} N_i^{\,2} p_i (1-p_i)/w_i}{N^2 d^2}$$

where $w_i = \frac{n_i}{n}$ or the fraction of observations allocated to stratum i. If equal allocation is used, $w_i = \frac{1}{L}$ for all strata. If proportional allocation is used, then $w_i = \frac{N_i}{N}$ for the ith stratum.

If equal allocation is used, then the final estimates of overall prevalence have to be weighted according to the stratum size. If proportional allocation is used, then the final estimates of prevalence do not have to be weighted: that is, proportional allocation is usually the preferred method because this simplifies subsequent statistical analysis of data.

Example: a survey is desired in three geographical areas, A, B and C, with population sizes of 3000, 5000 and 2000. The estimated prevalence rates (pi) of HIV in these three areas

are 15%, 10% and 12%, respectively. The task is to determine the sample size required for stratified sampling in these areas, such that the prevalence of HIV can be estimated within 2% of the true population prevalence with 95% confidence. The sample is to be distributed using proportional allocation.

Area	N_i	w_i	N_i^2	P_i	$(1 - p_i)$	$N_i^2 p_i(1 - p_i)/w_i$
A	3 000	0.3	9 000 000	0.15	0.85	3 825 000
B	5 000	0.5	25 000 000	0.10	0.90	4 500 000
C	2 000	0.2	4 000 000	0.12	0.88	2 112 000
Total	10 000					10 437 000

Using the above formula, the sample size is calculated as

$n = ((1.96)^2 * 10\ 437\ 000)/((10\ 000)^2 * (0.02)^2)$

≈ 1003

With proportional allocation, the sample will be distributed among the three strata as follows:

$n_1 = (1003 * 3000)/10\ 000 = 301$

$n_2 = (1003 * 5000)/10\ 000 = 501$

$n_3 = (1003 * 2000)/10\ 000 = 201$

Cluster sampling is generally less efficient than a simple random sample of the same size, and more units have to be included in a cluster sample to obtain the same degree of precision as that of an unrestricted simple random sample (Som, 1973). The easiest approach to calculate the sample size for two-stage cluster sampling is to calculate n based on simple random sampling and then multiply by the design effect to obtain the required total sample with cluster sampling.

Thus, the required total sample size when using cluster sampling is approximately

$n_c = n$ DEFF

$= n(1+(m-1)\rho)$

where m is the number of elements in each cluster, ρ is the intracluster correlation coefficient and n is the sample size required for the given precision had a simple random sample been used (see section 2.4 for estimation of design effect).

For example, if the design effect is 2, then twice as many observations would be necessary with cluster sampling as with simple random sampling to obtain the desired level of precision. (This implies that, if the same number of observations were to be selected for cluster and simple random sampling, the variance with cluster sampling would be twice as large as with simple random sampling.) Population-based surveys typically use a design effect of about 2, and it varies between 1.3 and 4.2 for demographic and health surveys dealing with fertility and reproductive health variables (Thiam & Aliaga, 1999; Ruilin Ren, DHS Macro International, personal communication, 2005).

Population-based surveys generally have to take the response rate (r) into account when estimating the sample size. The sample size for estimating p using cluster sampling, taking into account the design effect and the response rate, will then be:

$$n = \frac{z^2 p(1-p)}{d^2} \frac{DEFF}{r}$$

Table 2 shows the required sample sizes[4] for different levels of assumed prevalence (p) and precision (d) when the design effect is equal to 2 and the response rate is estimated to be 0.72. (For the response rate r, we assumed a household response rate of 90% and an HIV test response rate of 80%, hence $r = 0.9*0.8 = 0.72$.)

For example, for 5% prevalence and 2% precision,

$$n = \frac{(1.96)^2 (0.05)(0.95)}{(0.02)^2} \times \frac{2}{0.72} = 1267$$

Table 2. Sample size for a given HIV prevalence and level of precision

Precision[1]	Prevalence							
	3%	4%	5%	10%	15%	20%	25%	30%
±1%	3 105	4 098	5 069	9 604	13 606	17 074	20 008	22 409
±2%	776	1 024	1 267	2 401	3 401	4 268	5 002	5 602
±3%	345	455	563	1 067	1 512	1 897	2 223	2 490
±4%		256	317	600	850	1 067	1 251	1 401
±5%			203	384	544	683	800	896

[1]Precision is related to the width of the confidence interval.

Table 2 shows the minimum sample size required if the survey has only one sampling domain. If the HIV prevalence needs to be known to a given precision for different domains or subpopulations in a country (such as region or province), the required sample size has to be estimated for each of these subpopulations. The total sample size will then be the sum of the sample sizes for the different subpopulations.

The sampling has implications for the statistical analysis of data, and the sampling strategy should therefore be designed in consultation with a professional statistician.

2.6. Data to be included in the survey

Given the considerable expense of population-based surveys and the interest in behavioural and monitoring data, population-based surveys designed to measure HIV prevalence should also aim to collect information on social, behavioural and biomedical factors.

Box 2 provides a list of recommended variables for inclusion in HIV seroprevalence surveys (UNAIDS/WHO, 2003a). Where possible, indicators constructed from these variables should conform to internationally adapted indicators. Surveys should collect information related to an individual's participation in HIV testing, place of residence, demographic characteristics (age, sex and education), health status, exposure to HIV programmes, care and support and, if required, on HIV/AIDS-related knowledge, attitudes and behaviour. In addition, questions related to women's reproductive health and recent fertility history are recommended

to be included to enable better reconciliation of findings with sentinel surveillance, especially antenatal clinic surveillance. Questions related to use of contraceptives could provide useful information on condom use and reproductive health behaviour.

Including the line number of the respondent's spouse or cohabiting partner (if present in the household) will allow analysis of HIV concordance and estimates of assortative behavioural matching within couples. Information on marital status is much more useful if married people are categorized by whether they are remarried or in their first marriage and whether the marriage is monogamous or polygamous.

Box 2. HIV-related data to include in a survey

HIV testing
- Consented to HIV testing in survey
- Tested for HIV

Place of residence
- Region, province and district
- Urban or rural
- Urban, town, village or rural
- Recent migration

Demographic
- Age
- Sex
- Marital status
- Education
- Occupation
- Line number of spouse if present in household

Health status
- Knowledge of their own and partner's HIV status
- Circumcision status
- Contraceptive use
- Self-reported symptoms of sexually transmitted infections (such as genital discharge or genital ulcer)

Exposure to programmes
- Use of mother-to-child transmission programmes
- Use of voluntary counseling and testing programmes
- Exposure to mass-media messages

Care and support (appropriate only in countries with high HIV prevalence)
- Care and support for chronically ill people in household
- Presence of orphans in household
- Care and support for orphans

Health service utilization

Reproductive health (women only)
- Pregnancy status
- Date and survival of last live birth
- Use of antenatal care facilities in last pregnancy

Behaviour

- Psychosocial determinants of risk behaviour (such as risk perception)
- Health-seeking behaviour for sexually transmitted infections
- History of injection use
- Sexual behaviour (marital status and cohabitation, age at first sexual intercourse, number of partners, characteristics of partnerships, frequency of sex, condom use and contacts with sex workers)
- Other risk behaviour (injecting drug use, men who have sex with men, etc.)

Knowledge

- Knowledge of modes of transmission
- Sources of knowledge of HIV/AIDS
- Misconceptions about transmission
- Knowledge of ways to reduce risk of transmission
- Knowledge of other sexually transmitted infections

Attitudes

- Attitudes towards HIV/AIDS, including stigma and discrimination

For the individuals who are absent during the time of the interview, interviewers should still try to collect some basic sociodemographic information. This can be done as part of the household questionnaire. In addition to collecting information on household characteristics, information can be collected on all family members in the household including age, sex, marital status, residence (urban or rural), employment, and mobility (e.g., whether the family member lives away from home, for how long they have been away during the past year, where they are residing, and how frequently they come home). This information will be important when analysing data from nonrespondents to determine whether they differ in any way from the respondents in the survey.

Testing for other biomarkers

In addition to the standard demographic, HIV and behavioural data collected for the survey, considering other biomarkers for inclusion in the survey may also be of interest, such as tests for anemia, syphilis or other sexually transmitted infections.

Anemia testing has become a standard part of the demographic and health survey core household questionnaire in some countries (such as where malaria is prevalent) (ORC Macro, 2003). Collecting different types of biological samples and performing multiple tests will, however, add to the complexity of the survey process (for example, more training and supervision will be needed to collect specimens properly) and could increase refusal rates.

Separate informed consent must be obtained for each biological sample to be tested. If informed consent statements for different tests are combined, key elements could be missed (Yoder & Konaté, 2002). Surveys with multiple tests should therefore include additional testing, training and supervision of informed consent to ensure that all key elements for each test are included in the informed consent statements.

If tests are performed for treatable illnesses (such as sexually transmitted infections), the ethical implications of treatment for these conditions should be considered. Ethical standards require that counselling and treatment be given in cases where individuals receive results from rapid tests. One-time treatment may be feasible for certain illnesses, such as syphilis, but providing long-term treatment could be beyond the means of a survey.

2.7. Designing questionnaires

Once decisions have been made in relation to the data that will be collected, the questionnaires can be designed. The questionnaires should include a questionnaire number, questions on geographical and demographic information, as well as questions related to health status, reproductive health, care and support, service utilization, knowledge and behaviour, as listed in Box 2. HIV test results should not be recorded on the general questionnaire but entered into a separate file. Linking the demographic and behavioural data with the HIV test results, while ensuring confidentially and anonymity, is discussed in section 4.4. Examples of questionnaires used in demographic and health surveys are available at the MEASURE DHS+ (2005a) web site.

2.8. Strategy for HIV testing

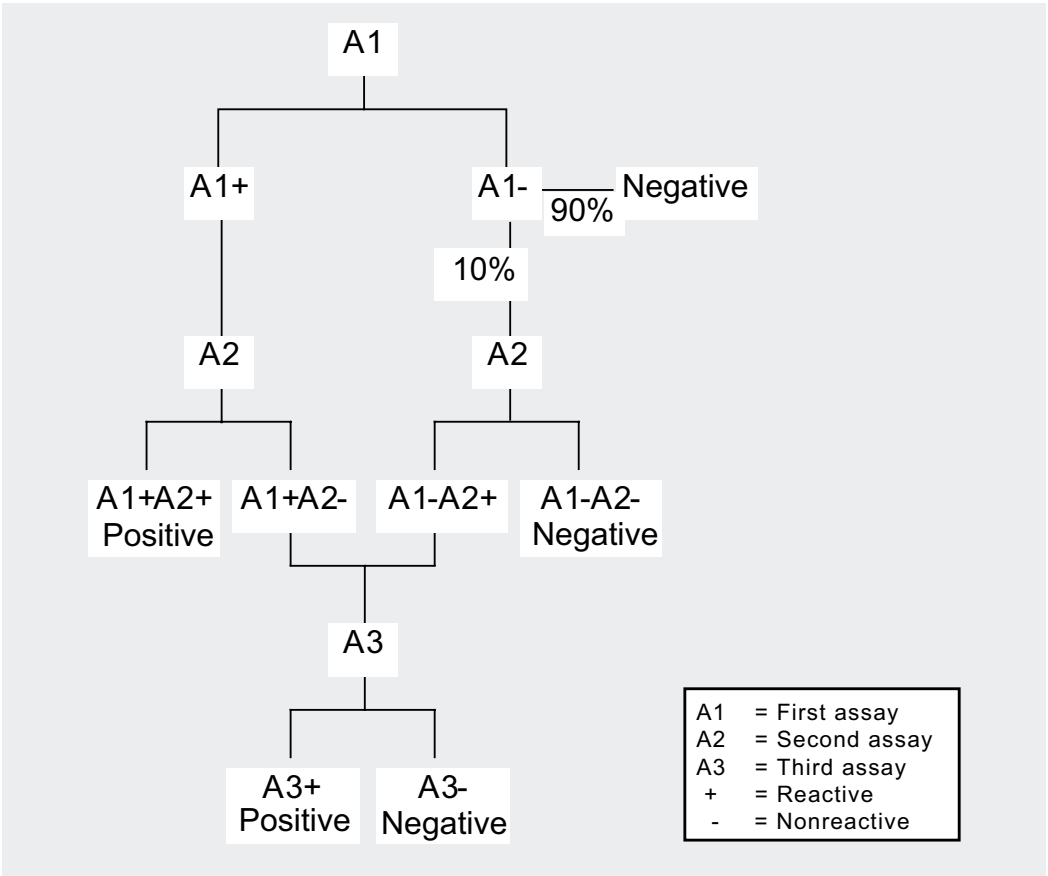
Providing estimates of HIV prevalence will be a core objective of the proposed population-based surveys, and obtaining HIV test results that are accurate and reliable is essential. Decisions have to be made in relation to the type of specimens to take and the type of tests to conduct on those specimens. HIV tests exist for blood, saliva and urine. Generally, surveys use blood or saliva samples to test for HIV (Table 1). Oral fluid and urine tests may appear more appealing because they are noninvasive and may present less biological safety concerns than blood collection but are currently not recommended because they lack many of the advantages (ability to distinguish HIV-1 from HIV-2, identifying viral subtypes, establishing antiretroviral drug resistance and application of detuned enzyme-linked immunosorbent assay (ELISA)) offered by blood-based methods (Respess et al., 2001). Further, the results from these type of specimens cannot be confirmed by a second test. For survey purposes, where testing is done at a central laboratory, the recommended type of sample is a dried blood spot, where a small amount of blood is collected from a finger prick on filter paper. Once dry, these samples can be stored at room temperatures for up to 30 days and can be analysed to the same extent as serum or plasma (for subtype determination or molecular studies) (Respess et al., 2001; WHO/UNAIDS, 2004b). The dried blood spot technique also makes collecting and transporting samples easier. All laboratory tests, however, can result in false-positive and false-negative results, and UNAIDS, WHO and the United States Centers for Disease Control and Prevention have therefore given guidance on conducting multiple confirmatory tests (UNAIDS/WHO, 1999: UNAIDS/WHO Working Group on Surveillance, 2001).

HIV testing strategies have been devised by UNAIDS and WHO to maximize the sensitivity and specificity of HIV tests while minimizing costs (UNAIDS/WHO, 1999; UNAIDS/WHO Working Group on Surveillance, 2001; WHO/UNAIDS, 2004b). The number and types of tests recommended in these strategies depend on the reason for the test (surveillance, blood screening or diagnosis) and the level of prevalence in a country. In agreement with proposed quality assurance for HIV testing and in order to make testing consistent over time and across countries, UNAIDS and WHO recommend that a second confirmatory test be conducted for all cases in which the first test for HIV is positive. In countries where the prevalence of HIV exceeds 10%, UNAIDS and WHO also recommend doing a second test on 10% of all cases where the first test shows a negative result (UNAIDS, 2004b). The strategy presented in Fig. 2 adds a third test as a confirmatory test when the first two tests are discordant (not in agreement). In countries where HIV-2 exists, the second test should be able to discriminate between HIV-1 and HIV-2. If the first two tests are not in agreement and a third test is not performed, then the HIV test result will be regarded as indeterminate.

The preferred HIV testing assay will depend on the purpose of the test (whether it will be used for surveillance purposes only or whether test results will be provided to individuals) as well as on the capabilities of the setting in which testing is being carried out. ELISA, which

has high sensitivity and specificity, is usually preferred where large numbers of specimens are to be run and rapid turn-around is not required. Rapid tests, however, may be preferred when small numbers of samples are being tested, limited facilities are available and a short turn-around time to provide a result is needed. Although rapid tests are technically simpler to perform and require less laboratory facilities and equipment, they can be more labour-intensive for the technician (Respess et al., 2001). HIV testing methods are evolving quickly, and staying abreast of new developments is important, particularly in terms of rapid tests. The status and characteristics of currently available rapid test kits may have to be compared to find the best suitable test, although many of them have sensitivity and specificity comparable to ELISA (WHO/UNAIDS, 2004b). The commonly used rapid anti-HIV antibody tests are based on the principles of dot immunoassay or particle agglutination (such as gelatin or latex). The general WHO and UNAIDS recommendation for countries is to use ELISA and/or rapid testing strategies for HIV antibody detection as well as for confirmatory testing. WHO and UNAIDS provide more detailed information on the characteristics of specific HIV assays (UNAIDS/WHO, 1999, 2002a, 2002b; WHO/UNAIDS, 2004b), including on the internet (WHO, 2005a).

Fig 2. HIV testing strategy



Quality assurance

The accuracy of test results requires high-quality work in obtaining the biological specimens, storing, tracking and transporting specimens and conducting the HIV tests in a quality-assured laboratory. All laboratories and testing sites carrying out HIV tests should have a well-functioning quality management programme. Quality control and assurance procedures must be strictly complied with to maximize the accuracy of the laboratory results. Procedures for detecting both laboratory and clerical errors must be included in the proposal.

Safety

In addition to ensuring the accuracy of HIV tests, specimens must be collected and tested safely to minimize occupational risk. The risk of harm to those being tested and those performing the test, as well as any exposure to biological hazard, must be minimized at all times. More information on laboratory biosafety and transport of infectious substances can be found in the *Laboratory biosafety manual* (WHO, 2003) and on the Communicable Disease Surveillance and Response section of the WHO (2005b) web site.

2.9. Budget planning

Resource and financial constraints in a country may limit the survey design. A carefully constructed budget will help determine whether sufficient human and financial resources are available to implement the survey. In addition to showing the necessary financial resources required for each stage of the survey, the budget should show the human resources required in terms of the number of personnel for each task and the time requirements for the personnel. Annex 1 shows an example of a budget with major groupings and items within groupings (although items may vary between surveys).

In determining the budget, special consideration should be given to the costs of HIV testing, such as the cost of hiring staff to take HIV samples and the related costs for training and supervision. It is often recommended to train interviewers to take blood samples, as this may reduce nonresponse (UNAIDS, 2004b), but this may be constrained by laws in the country about who is allowed to draw blood samples.

The provision of mobile voluntary counselling and testing services could significantly add to the cost because additional staff, vehicles, assays and other laboratory supplies are required (logistically it will be complicated if counsellors are required to stay behind the main survey team to provide post-test counselling). These costs may increase the overall survey costs substantially. The less expensive alternative is to provide referral to the nearest existing voluntary counselling and testing centre, assuming that such a system is available and functioning in the area.

2.10. Ethical considerations

Ethical issues to consider when including human subjects in a survey are guided by four principles. Study participants and associated communities should:

- be protected from any harm;
- participate in the benefits of the research;
- be informed of the procedures and risks; and
- freely choose whether or not to participate in the study.

The first principle is to protect the survey participants and the community from any harm, including stigma associated with HIV infection. The HIV status of every participant should be kept confidential, and there should not be any chance of the result being made public. Special safeguards are also necessary to protect all participants from any harm related to testing procedures.

The second principle (to participate in the benefits of the research) has, in the past, been handled in a general way in survey research, and surveys are often considered justified based on the societal benefits. The inclusion of HIV testing procedures in surveys has made more prominent the issue of how communities and individuals will benefit from

research. Although the communities involved may still only receive indirect benefits from improved programmes at national and regional level, individuals should benefit directly from participating in a survey. For example, individuals participating in HIV-related surveys should have access to counselling and testing facilities. In order to protect the confidentiality of the individual, the survey might not be structured to provide test results directly to individuals, in which case referral to free voluntary counselling and testing should be part of the survey protocol.

The third principle requires informing the study participants and community of the study procedures and potential risks involved. For non-medical personnel who will take specimens in the field, the procedures should be kept simple: for example, dried blot spots is a simple and safe means to obtain blood specimens for testing. Communities and community leaders or representatives should be consulted about the survey before the survey is conducted in a community. Informing community leaders of the purpose of the survey before conducting the survey has become a normal part of survey research. However, surveys with HIV testing will require additional discussions with community leaders to properly inform them of the procedures and the implications of the tests.

The fourth principle emphasizes the voluntary nature of participation in the survey. Individuals should freely choose whether or not to participate in research and should not in any way be coerced into the test by promises of benefits such as free treatment.

Extensive reviews and reports on the ethics of research with human subjects and of the ethics of HIV research have been written (Love, Thomson & Royal, 1999; MacQueen & Sugarman, 2003; Office for Human Research Protections, Department of Health and Human Services, 2004; WHO/UNAIDS, 2004a). Ethical principles and guidelines related to survey participation and the protection of human subjects as outlined in the Belmont Report (respect for persons, beneficence and justice) (Department of Health, Education and Welfare, 1979) and WHO ethical guidelines for consideration in second generation surveillance (WHO/UNAIDS, 2004) may provide additional guidance when considering the ethics of HIV testing in surveys.

Receiving ethical approval for a survey

To ensure that the above ethical principles and any country-specific ethical principles are adhered to in surveys, the procedures for the proposed studies must undergo review by appropriate ethical review boards. Where human participants are involved, review of the proposed study and ethical approval for conducting the study has become a requirement. In addition to the ethical review committee in the country where the survey will be conducted, other (international) bodies may need to review the survey protocol if organizations outside the country of the survey are significantly involved in the design and conduct of the survey. For example, demographic and health surveys have an institutional review board at ORC Macro International (the organization in charge of MEASURE DHS+) and an ethical review committee in the country of the survey to provide ethical clearance for the study.

The overall objective of these boards is to review survey procedures to ensure that the proposed study adheres to ethical principles. In addition to performing tasks normally associated with an ethical review, the board will review:

- informed consent forms and procedures to ensure that participants are fully informed of the procedure and potential risks and that participation will be voluntary;
- plans for handling of specimens to ensure that procedures can be implemented safely and effectively in the field;
- procedures to protect the anonymity and confidentiality of tested individuals;

- the risks and benefits to vulnerable or marginalized groups; and
- the provisions for participants to access voluntary counselling and testing.

2.11. Designing informed consent procedures

An essential requirement when biological samples are collected from survey participants is that they be properly informed of all tests to be conducted before choosing to participate. A well-designed informed consent form should include a description of the procedures and the implications thereof in the respondents' own language that is easy enough for all participants to understand. The purpose of informed consent is to inform individuals of all the procedures and any potential risks involved and to allow them to freely decide whether or not to participate in the survey.

Informed consent should firstly be obtained for participation in the survey, and separately for provision of biological samples to be tested for HIV or other infections.

For participants to be truly informed, they must understand the implications of the consent. However, ascertaining whether the individual really understands the implications of consent is difficult, as shown in a demographic and health survey conducted in Mali in 2001 (Yoder & Konaté, 2002). One option is for survey staff to pose questions to the individuals; another would be to allow the survey participants to pose questions to the interviewers. Allowing individuals to ask questions will help clarify the process and could increase the response rate.

Box 3 lists key elements in an informed consent statement. In addition, provision of test results or referral to a voluntary counseling and testing centre should be explained in the informed consent statement. Survey staff must be trained to include all these elements in the informed consent statement, especially if the informed consent statement is not recited verbatim or if the statement needs to be translated into a local language or dialect. Careful pretesting and training is necessary to ensure that the interviewers cover the key elements.

Annex 2 provides an example of an informed consent statement for a) participation in a survey and b) for HIV testing, which can be adapted to local laws and social norms.

Box 3. Key elements in an informed consent statement

1. Explanation of the survey purpose
2. Description of the procedures
3. Explanation of the risks and benefits
4. Description of how anonymity and/or confidentiality will be protected
5. An opportunity for the participant to ask questions
6. Statement that participation is voluntary and refusal will not affect any potential benefits

2.12. Planning survey operations

Staff organization

Conducting a high-quality survey requires well-trained and well-organized staff. Although the entire survey process from design to issuing a final report may take several months (12 to 15 months or longer if the survey covers a large sample), the workload varies greatly over the duration of a survey, with the most labour-intensive period occurring during

the 2–6 months of fieldwork, data collection and data entry. Although good management is important throughout the process, it is especially important during the phases of intensive fieldwork and data entry.

The fieldwork and data entry phases should be well structured because the workload is intense. Field staff should be organized into teams, each of which has a supervisor, field editor, male and female interviewers and, where needed, a technician to take HIV specimens for testing. Annex 3 provides an example of the staff required for the survey and brief descriptions of tasks. Each team should be able to operate independently, with the team leader assigning tasks, handling logistics and maintaining quality control. The survey director or survey specialists or consultants often assign enumeration areas to specific teams. Large surveys could have several survey specialists in the field during the fieldwork to provide general oversight and quality control, in addition to assigning enumeration areas to the teams. Although higher-level management should be available to assist team leaders, the team leaders should be capable of running the teams independently with only occasional oversight from the survey specialists or consultants.

The field interviewers should ideally collect HIV specimens, because they will have gained rapport with the respondent during the interview. Moreover, having the interviewers collect the specimens will reduce the logistical problems caused by having only one person in the team take the biological specimens. For example, if only one team member takes specimens, that person will have more difficulty scheduling callbacks due to the workload, which could result in a lower response rate for providing HIV samples. A potential difficulty with having field interviewers take the specimens is their lack of medical knowledge, and they will therefore require special training to take biological specimens.

The number of teams and size of teams depend on several factors, including the sample size and number of days that have been scheduled for fieldwork. Teams will generally only have one team leader, but the number of interviewers can vary. Teams typically have between four and six interviewers, including female as well as male interviewers. A simple way to calculate the required number of interviewers is to divide the sample size by the number of days scheduled for fieldwork and divide that by the average number of interviews done per day. For a demographic and health survey in which biomarkers are included, an interviewer typically conducts 2–3 interviews per day. For example, a sample of 4500 people could be done by 50 interviewers in 30 days (4500 interviews/(30 days times 3 interviews per day)) = 50 interviewers). The number of teams can then be obtained by dividing the total number of interviewers by the number of interviewers per team: 50 interviewers/(4 interviewers per team) ≈ 13 teams. A total of 52 interviewers would be selected for the survey to make sure each team has 4 interviewers.

As with fieldwork, data management should be structured and well planned. The data manager usually oversees the entire operation related to data management, cleaning and entry. Data entry specialists are usually responsible for writing the data entry programmes and data cleaning. Office editors usually track the questionnaires, edit the questionnaires and code the responses where required. One data entry specialist and one office editor are usually sufficient. Data encoders usually enter the data into the database, and the number of encoders depends on the number and size of the questionnaires. Ideally, data should be entered twice by two different data encoders to minimize encoding errors. Section 4.3 discusses data management in more detail.

Staff involved in designing the survey, analysing data and writing the final report usually include survey specialists and consultants. How they organize themselves depends on their skills and the schedule for completion of these tasks.

Mobilization

When a survey is conducted, mobilizing those who will assist with the survey and those participating in the survey is important, including the national and local government, research organizations and communities. In order to be most effective, mobilization should begin during the survey planning stage and continue through the survey implementation process.

The purpose of mobilization is to provide information about the objective of the survey, its design, implementation, utilization of the survey findings, the need for collaborative effort and the need for community participation. Mobilization should include members of the cabinet, parliamentarians, senior officials of key ministries, provincial and district administrative authorities including health officers, local-level government officials, religious and cultural leaders, international and local nongovernmental organizations, community-based organizations and organized groups or associations such as networks of people living with HIV/AIDS, youth groups and women's groups.

Mobilization themes and messages should focus on the type of audience and the delivery of messages and can be communicated through radio, television and other electronic and print mass media, formal letters to stakeholders and interpersonal interaction (meetings and personal discussions).

At all levels from national to local, issues related to HIV testing need to be explained, including confidentiality and anonymity and the benefits of the survey to those involved.

At the local level, mobilization is crucial for conducting the survey. Local government officials should be mobilized to assist with the survey, before the survey starts as well as during the survey. These officials, in turn, assist in mobilizing community members to participate in the survey.

2.13. Preparing manuals and pretesting the questionnaire and HIV testing procedures

In the final phase of planning the survey, the interviewer and supervision manuals should be written and tested. The interviewer manual will provide guidance on how to conduct interviews and collect biological specimens. The supervision manual will provide guidance on supervising teams, maintaining quality assurance and tracking questionnaires and specimens.

Before the survey starts, the questionnaires, specimen collection and HIV testing procedures should be tested. Questionnaires will be tested for the clarity of the questions, the flow of the questions and the time required to complete the questionnaires. Procedures for collecting and testing specimens will be pretested to assess: 1) the acceptability of the test to the population surveyed; and 2) whether logistics are in place for specimen collection, transport, verification and storage.

The pretest will reveal any problems in the design of the questionnaires or procedures, which will allow for corrections before the fieldwork starts. Once the pilot study has been conducted and any problems have been resolved, all survey staff are trained.

2.14. Training staff

All personnel involved in the survey must be trained to conduct surveys with adequate quality and oversight. After appropriate individuals have been identified and recruited at the local, regional and national levels, survey staff, including interviewers, supervisors, technicians,

laboratory staff and data management staff (office editors and data encoders) should undergo intensive training before the survey starts. The training must provide an overview of the HIV/AIDS epidemic in the country, including a discussion of HIV transmission, prevention, care and support. Training should review the survey objectives, operational procedures and field protocol. Issues related to ethics and confidentiality, voluntary anonymous testing, data and specimen collection, informed consent, survey management, laboratory testing and data management should be addressed. Taking specimens for HIV testing should be explained and demonstrated. During the training sessions, survey staff should have the opportunity to discuss concerns and obtain clarification on survey operations or to share any previous experiences. All staff should feel confident with the survey procedures in which they will be involved.

The purpose of training is to ensure that interviewers, supervisors and data entry personnel clearly understand their roles and responsibilities and know how to carry out their tasks before the fieldwork starts. In addition to thoroughly understanding the questionnaires, the interviewers (or people responsible for collection of biological specimens) need to understand the procedures related to obtaining consent and must be able to collect biological specimens. Supervisors must know how to ensure the quality of the processes of interviewing and specimen collection, manage field logistics, track questionnaires and biological specimens and determine work assignments. Data entry personnel must understand the data entry process and how to enter questionnaire data into computers, check for discrepancies in the data and correct data entry errors. Data editors need to know how to ensure the quality of the information on the questionnaires and track questionnaires if errors need to be corrected.

Maintaining motivation among survey staff will facilitate high-quality work and completion of all survey activities. Motivation of staff can be maintained by:

- developing a sense of survey ownership among staff;
- clearly defining the responsibilities and roles of all staff at all levels;
- emphasizing the importance of each person's contribution;
- providing adequate staff training;
- ensuring that necessary equipment is available to conduct the survey; and
- providing feedback on staff performance, statistics on progress and provisional survey results.

3. Including measurement of HIV prevalence in existing national population-based surveys

HIV testing can be included as part of existing population-based surveys (such as demographic and health surveys, AIDS indicator surveys and multiple indicator cluster surveys) providing that these surveys have a sufficiently large sample size and are representative of the populations of interest. This usually implies that the survey should be representative at a national and regional level and of urban and rural areas.

An important consideration when adding HIV testing to an existing survey proposal is that it will increase the cost and complexity of the survey. However, this is generally less expensive than designing an independent HIV seroprevalence survey despite the added cost. A further concern is that the inclusion of HIV testing might affect participation rates. Individuals may decide not to participate in the main survey because they know that HIV testing will be part of the survey.

Including HIV testing in existing survey protocols involves steps similar to those for designing a new survey. However, some of the steps may be simpler because the study design might already have been done for the original survey and a protocol might already exist. The following main issues in relation to measuring HIV should be considered in the planning phase.

1. Identify the objectives of the survey, specifically in relation to HIV. Secondary objectives might be similar to those listed in section 2.1. Decisions have to be made as to whether to include certain subgroups for HIV testing (such as children younger than 15 years of age).
2. Check the sampling methods to ensure that the sample size is representative of the study population and that sample sizes are sufficiently large to obtain reliable estimates of HIV for various subgroups or strata (such as urban or rural and provinces or regions). If the sample size of the existing survey is not sufficient, it may have to be increased to ensure reliable estimates of HIV prevalence (see sections 2.4 and 2.5).
3. Decide on the HIV strategy: in particular, whether blood or saliva are going to be used, the type of test and assays, the procedures for confirmatory testing and quality assurance for the laboratory. The laboratory that will perform the HIV tests has to be identified. Section 2.8 describes the recommendations, which are the same as for designing a new prevalence survey.
4. Decide whether to collect any additional data in relation to HIV such as data on sexual behaviour and knowledge and ensure that relevant changes are made to the questionnaires. Box 2 provides a list of variables recommended to be included in HIV prevalence surveys. Ensure that all demographic, geographical or other information required for specific subgroup analysis is included in the questionnaire. Decisions also have to be made on whether to collect data on other biomarkers, such as whether to test for syphilis, other sexually transmitted infections and anemia (section 2.6). In addition to indicating whether an individual agreed or refused to participate in the survey, the questionnaire should indicate whether the participant consented (or refused) to be tested for HIV. Whether a selected individual was absent during the survey day should also be noted.

5. Informed consent forms for participation in the survey as well as for providing a blood sample for HIV testing have to be designed, as described in section 2.11 (examples given in Annex 2).
6. Procedures for providing the study population with access to voluntary counselling and testing should be considered (section 4.2).
7. Procedures should be developed to ensure the confidentiality of participants at all stages during the survey (described in section 4.2) and during the management of data. Section 4.4 discusses linking HIV data with other demographic, behavioural and biomedical data such that the confidentiality and anonymity of participants are ensured.
8. Non-response can lead to potentially biased estimates of HIV and it is important to try to maximize participation rates (such as revisiting households in the case of absenteeism). Care should however be taken to ensure that an individual does not feel in any way coerced into participating or providing a blood sample. Individuals who do not want to provide a blood test should still be encouraged to participate in the main study because information collected from the questionnaire will be important for trying to understand and address nonresponse bias. Chapter 5 discusses potential bias related to nonresponse and ways to adjust for nonresponse.
9. Including HIV testing in an existing survey increases the overall cost. An existing budget has to be adjusted accordingly.
10. Ethical issues related to including HIV testing in the survey should be considered (section 2.10). These are the same as for conducting a new prevalence survey.
11. Ethical approval for the study needs to be obtained from an ethical review committee or institutional review board that reviews all survey procedures to ensure that the study proposal adheres to ethical principles (see section 2.10).
12. Survey staff have to be organized, mobilized and trained, and questionnaires and HIV testing procedures have to be pretested. If the HIV testing is added to an existing survey, staff have to feel comfortable and confident about dealing with the HIV testing and additional questions in the questionnaire (discussed in more detail in sections 2.12–2.14).

4. Conducting a population-based survey to measure HIV prevalence

After procedures for collecting data and biological specimens have been pretested and survey staff have been trained, fieldwork will begin: collecting data through interviews and taking specimens for HIV testing.

4.1. Field operations

Field operations for a population-based survey usually start by collecting information on households in selected enumeration areas through household listings. The household listing provides a list of addresses of all households in the enumeration area and serves as a sampling frame from which households can be randomly selected for inclusion in the survey. This will also help the interview team in locating the selected households when visiting the enumeration area. In addition, the household listing should provide information on population size in each household in the specific enumeration area and hence in each enumeration area.

The primary aim of fieldwork is to collect high-quality information and biological samples from individuals while adhering to the sampling design. Collecting information from the correct enumeration areas and selected households will ensure a representative sample.

Field supervision is necessary to ensure that fieldwork is conducted in accordance to the survey proposal and to ensure high-quality collection of data and specimens. Team leaders are responsible for managing team member work assignments and field logistics and for ensuring quality control.

Further information on the general fieldwork can be obtained from the model manuals on how to implement demographic and health surveys (MEASURE DHS+, 2005a).

4.2. HIV testing

4.2.1. Collecting, handling, processing and tracking specimens

The people collecting biological specimens should follow routine procedures for the type of specimens that are being collected. For example, Annex 4 describes the procedures for collecting dried blood spots. Field staff who will be collecting specimens require special training. Safety as well as accuracy should be emphasized when collecting specimens. The United States Centers for Disease Control and Prevention (1988) have provided guidelines on universal precautions and safety in blood sampling. Annex 5 provides more information on handling and disposing of biohazardous waste.

If HIV results are to be linked with demographic and behavioural data, biological specimens should be clearly labelled with a survey code (such as a barcode) while ensuring the confidentiality of the participants. The survey code is a unique identification number assigned at the time the specimen is collected, which will be used to link the test results to other survey data. Storage of specimens also depends on the type of specimen. If serum samples are to be collected and tested for HIV more than three days after collection, they should be stored at –20°C. For longer term storage, serum should ideally be frozen at –70°C. Dried blood spot samples can be stored in storage bags at room temperature for up to 30 days or at 4°C for up to 90 days (George et al., 1989). In all cases, high humidity and very high temperatures should

be avoided. If dried blood spot samples are kept for more than 90 days, they should be stored at –20°C.

Specimens not tested on site need be transported to a regional or national laboratory for testing. Transport methods depend on the country's infrastructure, but field staff involved in the survey are frequently responsible for transporting the specimens to the nearest proposed laboratory. Dried blood spots can be transported without coolers, whereas serum specimens should be packed in coolers at 4°C.

Biological specimens need to be tracked to minimize loss of specimens in the field, during transport and during the testing phase. Each specimen should be labelled with the survey code, which will be logged onto the field laboratory report form. Each batch of specimens being transported to the laboratory should have a specimen shipment form indicating the number of specimens in the batch, the date and time of shipment, the codes of the originating enumeration areas, districts or regions and the name and signature of the team leader who will check specimens against completed questionnaires before being dispatched. Each specimen delivered at the laboratory will be logged in a laboratory register. Unique laboratory numbers might be assigned against which testing and storing in the repository will be carried out. Specimens received at the laboratory should be checked against the field laboratory report form to confirm the number of specimens sent and received.

For both linked and unlinked anonymous testing, no personal identifier should be included on the samples or in the reports apart from the survey code, which will be used to link the test results to other survey data. For linked confidential testing, the survey codes will be linked to personal identifiers to facilitate the provision of test results. Strict measures must be followed to ensure that such information is accessible only to laboratory and survey staff. After the participant has received the test result, all personal identifying information should be destroyed. For the purpose of future data analysis, HIV test results should be linked only by survey codes.

4.2.2. Ensuring confidentiality and anonymity

Survey staff should ensure that the confidentiality of all personal information collected from participants during the survey is protected. This is critical, and although all survey staff are responsible for this, the survey coordinator is ultimately responsible. Special care must be taken in the survey process to prevent the release of any information related to the HIV status of any participant. For serosurveys using linked HIV testing, mechanisms that ensure anonymity should also be used to ensure that test results are confidential. Section 4.4 discusses this in more detail.

Rules to protect participant anonymity include the following.

- The staff taking the biological specimens should not be performing the laboratory tests.
- Databases, laptops and forms should be protected. Databases must be password-protected so that only people working in the survey programme will have access. Laptop computers and survey forms should be locked in filing cabinets or drawers when not being used by survey staff.
- Access to HIV test results should be restricted. Data management staff should ensure that HIV test results are not available to staff who collected the demographic and behavioural data and/or biological specimens.
- The register of laboratory results and backup files should be locked in filing cabinets, and only authorized people should have access to them.

- HIV data should not be linked to questionnaire data until all identifying information have been removed from these files. Until such time, questionnaire data must remain separate from laboratory data in password protected computers with limited access.

4.2.3. Providing test results and voluntary counselling and testing

Testing survey respondents for HIV infection raises several ethical questions, including whether the respondents should receive the test result or not. An important principle in either case is that study participants should share in the benefits of the research, so that if study participants are found to have an illness, they expect something to be done. The ability to meet this expectation depends on the severity of the illness and the available treatments. For sexually transmitted infections, such as syphilis, where there is a relatively simple and effective treatment, test results can be given in the field along with referral for appropriate treatment. However, HIV poses a much more difficult problem because of the nature of the disease. Because there is no cure, treatment is needed over the course of a lifetime, and treating HIV-positive respondents for life is usually beyond the financial means of any survey.

One way of addressing this issue is to ensure that all study participants have access to voluntary counselling and testing services, either directly through mobile voluntary counselling and testing services during the survey or through referrals to voluntary counselling and testing centres in the area. This will give participants the opportunity to have their HIV status determined while being counselled and to receive care if needed. Several surveys in which HIV testing has been included have used the approach of referring survey respondents to the health care system for free counselling and testing. For example, in the demographic and health survey in Mali in 2001, referral cards for free voluntary counselling and testing were given to all respondents (Cellule de Planification et de Statistique du Ministère de la Santé [Mali], Direction Nationale de Statistique et de l'Information & ORC Macro, 2002). Similarly, referral cards for free voluntary counselling and testing were given to study participants in the Nelson Mandela/Human Sciences Research Council Study of HIV/AIDS in South Africa in 2002 (Shisana et al., 2002).

Voluntary counselling and testing can also be provided alongside the survey through the use of mobile voluntary counselling and testing services. For example, in the Kenya demographic and health survey in 2003 (Kenya Central Bureau of Statistics, Kenya Ministry of Health & ORC Macro, 2004) a two-member voluntary counselling and testing team accompanied the main survey team in areas outside Nairobi. In addition to the HIV test being done for the survey, respondents could receive free voluntary counselling and testing from this mobile team. The voluntary counselling and testing team carried out counselling followed by rapid HIV tests. Additional counselling was done when the test results were given to the respondents. Respondents may prefer this approach over a referral card due to convenience and rapid results. Positive HIV test results have to be confirmed by a second rapid test. The downside of such an approach is the added costs and the increased logistical complexity to the survey.

If the survey design proposes to provide HIV test results to the survey respondents, the testing strategy may need to be modified. Although the sensitivity and specificity of most HIV testing algorithms exceed 99%, HIV testing for the purpose of surveillance need not be 100% specific. However, HIV testing used to diagnose HIV infection (to inform individuals of their HIV status) does need to approach 100% specificity. Further, positive HIV test results should always be confirmed before informing individuals of their test results, as described in Section 2.8 (WHO/UNAIDS, 2004a).

4.3. Data management

Data management should follow a series of steps to produce high-quality data on household and individual characteristics as well as on HIV prevalence. The databases should be managed from a central location, and a database manager should be appointed to take charge of the process.

The steps involved in data management include the following.

- Questionnaires should be checked, edited and coded in the field (1) to correct errors on the questionnaire, (2) to code special responses (such as from open-ended questions) to facilitate ease of data entry and analysis and (3) to report back to field workers on common errors made in the field to improve the fieldwork. Where errors were made on the questionnaire, going back to the household to obtain the correct information is sometimes necessary.
- A database should be created, and data entry, checking and editing programmes should be written for entering data from the questionnaires into the electronic database and to check information in the database for potential data errors. Programmes need to be written and tested at the pilot stage. These programmes can include acceptable ranges for variables (such as allowing only valid entries within a certain range for a particular variable) or checks for internal consistency, which will help with data cleaning.
- Questionnaires should be logged and tracked to ensure that planned interviews have been completed and to keep track and provide a count of the number of questionnaires that have been entered into the database. They need to be stored in an accessible manner in case they need to be referred to during data editing.
- Information recorded on the questionnaires will be entered into an electronic database. Data entry should be done concurrently and continuously during the survey as data are collected. Data should be entered twice, ideally by two independent data encoders, to minimize data entry errors. Once entered, the two data sets should be compared and discrepancies corrected.
- Once entered onto a computer, electronic data should be checked for errors and outlying values, and all inconsistencies should be corrected, so that data files accurately reflect the responses to the questions in the questionnaires. Information generated on the quality of the data from the questionnaires can also be fed back to the field teams in an attempt to improve the quality of the fieldwork. Frequency tables can be prepared for all variables to check for outliers (singly and in logical pairs). Variables related to each other can be compared for inconsistencies: for example, a respondent's age can be compared with the date of birth. Distributions and scatter diagrams of variables should be plotted and checked for plausibility. Decisions have to be made about whether to impute or eliminate clear mistakes or missing data or whether to simply flag the relevant fields as invalid.
- Files with recoded variables can be created to make analysis of data easier.
- HIV data should be managed separately (and data entered into a separate file) by the laboratory that performed the HIV tests. Once both data sets have been entered and cleaned and personal identifiers have been removed, the HIV data can be merged with the data from the general questionnaire using the special preassigned survey code (such as a barcode).

4.4. Unlinked versus linked test results

Three different approaches to HIV testing, summarized in Box 4, are commonly used in HIV surveillance and seroprevalence surveys: unlinked anonymous testing or linked anonymous testing with or without providing test results to respondents.

Box 4. Summary of the various HIV testing approaches used in population-based surveys

Unlinked anonymous testing (with informed consent)

- Testing of unlinked specimens collected solely for surveillance purposes
- No personal identifiers or names obtained
- No counselling required

Linked anonymous testing (with informed consent but no test results provided)

- Testing of samples linked to individual characteristics by unique survey code (such as a bar code)
- No personal identifiers or names obtained
- No counselling required
- Referral to voluntary counselling and testing or for independent testing

Linked anonymous testing (with informed consent and provision of test results)

- Testing of samples linked to the person by survey code (preferably a bar code)
- Pretest and post-test counselling required
- No personal identifiers or names obtained
- Specimen coded; code given to participant so that only he or she may obtain result

Unlinked anonymous HIV testing completely protects an individual's anonymity because it does not provide any information that would allow an individual with HIV to be identified. No information that could reveal the participant's identity (name or questionnaire number) is provided with the blood specimen or field notes. In some surveys, only a small amount of information is kept with the HIV sample to facilitate analysis of HIV prevalence by geographical location, age and sex.

The disadvantage of unlinked testing is that it only allows limited statistical analysis, and HIV status cannot be linked with the social, behavioural or other biomedical information collected through the interview. Essentially this is an extension of the type of surveillance that is currently done in antenatal clinics – except that in a household survey (as opposed to a clinic) there may be no reason for requesting a blood sample other than HIV testing, in which case informed consent for the test must be obtained.

Linked testing

The effects of behaviour (and other individual information) on HIV status can only be studied if the HIV test results can be linked to the behavioural data, which is one of the main strengths of including HIV testing in population surveys. Linking HIV status with participant information allows researchers to analyse the association between HIV status and knowledge, attitudes, behaviour and other information. The results can provide essential information that will help programme managers in designing and evaluating HIV/AIDS prevention programmes.

The UNAIDS Reference Group for Estimates, Modeling and Projections recommends that "Whenever possible, surveys should link biological data with other data while protecting the identity of the participants" (UNAIDS, 2004b).

Two different types of linking can be used: linking to personal characteristics and behavioural data without providing results to a particular individual and linking to feed back test results to the individual concerned. The latter strategy implies that full pretest and post-test counselling is provided as an integral part of data collection. The former strategy should refer participants who wish to know their HIV status to a counselling and testing service (which could also be offered by mobile counsellors attached to the survey team), but separate samples are used for survey research purposes and for diagnostic purposes.

One potential problem with linking results is that, if this is not done in a careful and responsible manner, the HIV status of individual survey respondents' risks being exposed. However, careful procedures have been devised to prevent possible disclosure of an individual's HIV status and have been implemented in several countries (Kenya Central Bureau of Statistics, Kenya Ministry of Health & ORC Macro, 2004; Sharman, 2000).

One way to prevent the test results from being linked to a name or enumeration area that could identify an individual is to have different organizations manage the HIV test and demographic information or have it managed at different locations. HIV test results should ideally be managed at the laboratory responsible for performing the HIV tests. Once the data have been entered onto computer and cleaned and any information that could identify an individual has been removed from the files, HIV data can be merged with the household or individual data using the unique survey code (such as a barcode).

An important question is how much geographical, demographic, household and individual information can be linked without enabling an individual to be identified. In general, if names and the location of small geographical areas are removed from data files, individuals cannot be identified. In large geographical areas, demographic information such as age and sex will not be sufficient to identify individuals.

The following information should not be linked with HIV test results.

- The name of the participant or any relatives or others living in the same household should never be linked with HIV test results. When an individual's name is recorded on the original household or individual questionnaire for the purpose of cross-checking information, the part of the questionnaire that contains the name must be destroyed after data cleaning and before the file is linked to HIV test results. Under no circumstances should names be entered into electronic databases (because ensuring that names are removed from all files and that a copy of the database with names has not been stored elsewhere may be difficult).
- Identification of a small geographical location (such as enumeration area code or the name of a small town or village) where the individual lives combined with limited demographic information could be enough to identify an individual. For example, in a population smaller than 1000, the location, age and sex could provide enough information to identify an individual. As a general rule, geographical information below the district level (such as enumeration area codes) should not be linked with HIV test results. Completely eliminating enumeration area information from files with HIV test results might not be desirable, because the statistical analysis may have to adjust for clustering at the enumeration area level. A recommendation is that the enumeration area number can be included in the linked file as long as this number cannot be associated with the actual geographical location of the enumeration area. It might therefore be necessary to recode enumeration areas in such a way that the new number cannot be associated with the geographical information for the enumeration area. For example, a new and unique number can be selected randomly for each enumeration area, and the enumeration area number for each individual in the file would be changed accordingly.

- A household number or a questionnaire number that can link the individual to the original questionnaire used in the interview should not be linked to HIV test results. (For merging files, the survey code (often a bar code number), and not the questionnaire number, should be used.) Household and questionnaire numbers that contain the geographical identification of the enumeration area should be removed from files. If the enumeration area code is embedded in the household or questionnaire number, it should be removed or scrambled at the same time that the enumeration area code is scrambled. Steps should be taken, however, to ensure that all forms that group together (for example, because they belong to the same household) before scrambling can still be grouped together after scrambling by electronically assigning another number that is unique within the enumeration area. Otherwise the ability to use household information (such as size, assets, sex and age of head of household) for individual-level analysis or mother information in analysis of child data could be lost.

Procedures for linking HIV test results with demographic and other survey information

Box 5 outlines the proposed steps in linking HIV test results with survey information. The assumptions are made that the HIV data will be managed separately (for example, by the laboratory performing the HIV test) from the organization that collects and enters the demographic and other data and that this latter organization will not have access to the HIV test results until the end of the process (that is, after all the data have been entered and cleaned and personal identifiers have been removed).

Box 5. Procedure for linking HIV test results with survey information

1. Obtain informed consent from participant for HIV test and take blood sample

2. Attach the unique survey code, such as a bar code, (essential for merging files) to:
   a) individual questionnaire and to consent form
   b) blood or saliva or dried blood spot card
   c) specimen tracking form

3. Analyse specimens in the laboratory and enter the HIV results separately into a database

4. Enter household listing, information from household questionnaires and individual questionnaires into data files, including enumeration area codes (ensuring that no name or geographical location information below district level is entered)

5. Clean data files and resolve any problems with location and identification information required for merging files

6. Remove names from paper questionnaires

7. Renumber enumeration areas such that they cannot be associated with geographical location

8. Data entry organization obtains test results from laboratory and merges these results with individual data files using survey codes

5. Obtaining estimates of national HIV prevalence from population-based surveys and existing sentinel surveillance

This chapter discusses potential biases related to sentinel surveillance, especially antenatal clinics, as well as population-based surveys and how to adjust for these biases in estimating national HIV prevalence. We discuss how population-based surveys (which are conducted less often but with better population coverage) can be used to improve sentinel surveillance-based estimates of HIV prevalence and associated trends. A description of UNAIDS/WHO methods for estimating and projecting national HIV prevalence from surveillance data is provided.

5.1. Calculating HIV seroprevalence

Basic assumptions for calculating HIV prevalence

Seroprevalence *(p)* can be calculated from survey data as

$$p = \sum_i \frac{y_i}{n}$$

where, at a given site or among a specific subgroup, $\sum y_i$ represents the total number of people who were positive for HIV out of the total number, n, who were tested. Multiplying this proportion by 100% will express HIV prevalence as the percentage positive.

For large enough ($n > 30$) sample sizes, 95% confidence intervals, based on binomial normal theory, can be calculated for p using the formula

$$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$

For small n ($n < 30$), confidence intervals based on exact binomial errors can be obtained using standard computer software. The confidence interval is interpreted as having a 95% probability of containing the true population prevalence.

Care should be taken when aggregating data from different survey sites, and important differences between sites or regions have to be taken into account. When national HIV prevalence is estimated from sentinel surveillance, care should be taken to ensure that the sentinel sites adequately reflect the overall national population.

Temporal HIV prevalence trends can be compared by site if standardized surveys have been repeated over time within the same sites using the same methods. Plotting HIV prevalence over time for specific sociodemographic groups (such as age groups) will help to identify and assess changes in levels of HIV infection.

5.2. Potential biases related to antenatal clinic surveillance and population-based surveys

Bias is a potential problem when HIV prevalence is estimated using either sentinel surveillance or population-based surveys. Bias associated with antenatal clinic data includes whether the pregnant women who attend public antenatal clinics are representative of all pregnant women, reduced fertility among HIV-infected women, selection for sexual activity and underrepresentation of smaller rural sites in surveillance systems (Boerma, Ghys & Walker, 2003). National population-based surveys are more representative of the general population than antenatal clinics and include women (pregnant and not pregnant) as well as men. Although these surveys are generally geographically representative, groups that might be at higher risk of infection (such as sex workers, migrant populations, army and police personnel, prisoners or others) may not be included in population-based surveys to the extent that their living arrangements (such as group quarters) are not covered as part of the household survey. In most instances one would expect estimates derived from population-based surveys to underestimate the true prevalence, especially in countries with relatively low prevalence where HIV is concentrated in groups with high-risk behaviour. However, the magnitude of this bias is likely to vary greatly between countries (UNAIDS/WHO, 2003a), depending both on the size of the high-risk groups that are not covered in the survey and the extent to which infection levels in the groups exceed the levels found in the general population. Special targeted surveys or surveillance efforts are needed if excluding these groups is considered to significantly affect the HIV prevalence estimates.

A potentially greater source of uncertainty in the estimates from population-based surveys is related to nonresponse of participants, either because they refuse to participate or are absent from the household. The sources of nonresponse and the procedures for adjusting estimates for nonresponse are discussed in detail in the sections below.

5.3. Nonresponse in population-based surveys

Nonresponse in population-based surveys (absence from the household or refusal to participate) is of concern because it can potentially bias the final HIV prevalence estimates. Where nonresponse is not associated with HIV, bias is not a problem, but where it is, HIV prevalence estimates will be biased. Nonresponse in population-based surveys could occur at the household level or at the individual level. At the household level, nonresponse may be related to the absence of all household members, or all adults, on the day of the survey, or heads of households could refuse to let household members be interviewed. Changes between the time of the household listing and the start of the fieldwork could lead to a dwelling not being found if it has been demolished or become vacant. The household response rate is calculated as the proportion of all households selected (excluding dwellings found to be vacant or destroyed) for which household interviews are completed (see Annex 6, model Table S1 as an example for recording household response rates).

Household response may relate to bias in HIV prevalence estimates in several ways:

- Absence of the entire household is more likely to be associated with small families, especially single-person households, and adults living in single-person households may be more likely to be infected with HIV.
- Absence may be associated with movement of household members to another location because of illness or recent death (Urassa et al., 2004).
- Household presence may be related to illness associated with HIV/AIDS, especially in rural areas where, for example, migrant workers may return home when they become ill.

Typically, household nonresponse in population-based surveys is low and detailed analysis may not be needed to estimate the effect on HIV prevalence. However, surveys with high household nonresponse rates are more likely to be biased than surveys with low nonresponse rates (Shisana et al., 2002).

Nonresponse at an individual level may be related to absence at the time of the test or refusal to participate. The reasons for absence from a household and refusal to participate differ, and hence the biases affecting estimates of HIV prevalence are likely to differ. Separating nonresponse due to absence and refusal is important in analysing the effects of nonresponse on HIV prevalence.

Refusing to be tested may be associated with a higher or lower risk of HIV infection compared with those who agree to be tested. Respondents may refuse testing because they already know their status or because of fear that they might be infected. Most studies, however, have shown that reported personal risk perception is weakly or moderately associated with HIV prevalence (UNAIDS/WHO, 2003a) and drawing general conclusions on the strength of the association between refusal and HIV prevalence is difficult.

There is evidence that frequent absence from a household is associated with increased HIV prevalence. People who travel more or families affected by labour migration have higher HIV prevalence rates than those who are not affected to the same extent by migration (Lydié et al., 2004; Zaba et al., 2004; Zuma et al., 2003). Short-term mobility (traders, businesspeople and people in search of work) is particularly important, and people making frequent short trips may not be available during the time the survey team visits the household. Survey teams are advised to make special efforts to improve response rates by making appointments and visiting households in the evenings or during weekends and by making follow-up visits. Typically in demographic and health surveys, interviewers are required to make at least three visits at different times and on different days.

In summary, cases where nonresponse may be associated with higher risk of HIV include single-person households, frequent travelling or labour migration, refusal because of fear of being HIV positive (implying risk) and refusal because of known HIV status (more likely to be HIV-positive). In contrast, people of higher socioeconomic status, higher education levels or older people are often unwilling to take part in surveys of HIV and may be at lower risk of HIV.

Assessing coverage of HIV testing

To estimate the coverage rates (and nonresponse) for HIV testing, information should be presented on testing status of individuals (percentage tested, refused testing and absent for testing and percentage missing data) by geographical area (urban or rural) and by sex and age group. For the above categories, the percentages interviewed and not interviewed are needed (see example in Annex 6, model Table S2). In addition, coverage can be reported by background characteristics, such as residence, region, marital status, high risk and sexual behaviour (as shown in Annex 6, model Table S3) to show how nonresponse varies for a given background characteristic. If there is little variation, nonresponse for that characteristic may produce little bias in HIV prevalence. On the other hand, the characteristics with high variation in nonresponse are likely to produce biased HIV prevalence estimates.

5.4. Analysis of nonresponse

Examining the variation in nonresponse in relation to the background characteristics of the population is important for trying to understand the impact of nonresponse on HIV

prevalence. If nonresponse is high for a group of people that is at high risk of HIV, prevalence estimates will be biased downward. Conversely, if nonresponse is high for a group with low HIV prevalence, estimates will be biased upwards. Characteristics showing variation in nonresponse that might affect HIV prevalence estimates include sex, geographical area (HIV prevalence is likely to be lower in rural areas), age (women 25–29 years old and men 30–35 years old have the highest prevalence rates), marital status (single people could be at higher risk if they have multiple partners, and widowed or remarried people have higher prevalence), socioeconomic status or sexual behaviour (for example, having a high number of sexual partners will be associated with high risk of HIV). To understand how these characteristics differ between those who refuse testing and those who do not, it is advisable to compare the proportions of these characteristics in the two groups in bivariate frequency tables. In addition, multiple logistic regression analysis can be performed (with presence or absence or testing status (yes or no) as the dependent variable for analysis of nonresponse due to absence or refusal, respectively) to analyse in more detail the relative contributions of each of the characteristics on nonresponse.

Performing statistical analysis to assess whether background characteristics differ between nonrespondents and respondents requires separating those who refused to be tested from those who were absent during the survey. If those who refused to be tested were willing to be interviewed, detailed analysis, including analysis of sexual behaviour, will be possible. For those who were absent during the survey, information on background characteristics will be limited to some demographic (such as age and sex) and household characteristics, and the analysis will be limited to these characteristics.

5.5. Adjusting survey estimates for nonresponse

The impact of nonresponse on HIV prevalence can be determined by comparing the differences in characteristics between those who were tested for HIV and those who were not tested, as summarised above. If nonresponse is likely to affect HIV prevalence estimates, decisions will then have to be made on how to adjust the national prevalence estimates.

Nonrespondents (especially those who refuse to be tested) do not always have a higher HIV prevalence than respondents, as is often assumed. For example, a study on the impact of nonresponse on HIV prevalence in Kenya found that the adjusted prevalence for women, using regression analysis, was slightly lower than the unadjusted prevalence (Johnson, Way & Otieno, forthcoming), whereas the adjusted and unadjusted prevalence rates for men were about the same.

Several approaches are available and have been used to adjust HIV prevalence estimates for nonresponse, two of which will be discussed below. The first is to standardize prevalence estimates based on characteristics known to affect HIV prevalence, and the second is to use regression analysis for those who were tested against their background characteristics to estimate the prevalence among those not tested (using the background characteristics).

Standardization approach

The standardization approach for adjusting HIV prevalence estimates for nonresponse involves calculating the expected number of HIV-positive people among those not tested. Information on the characteristics of nonrespondents is combined with the HIV prevalence among those who were tested to produce expected numbers of infections among those who were not tested (this is best done within the different categories of the characteristics being used in the adjustment). New estimates of prevalence are then obtained by adding the expected number of infections among nonrespondents to the known infections among those who were tested.

Adjusting for differences in HIV prevalence first requires specifying the desired assumptions about the relative risk (RR) of HIV among those who were not tested compared with those who were tested. For example, if RR = 2, the proportion of HIV infections expected among those who were not tested will be twice the proportion of HIV infections among those who were tested. Determining the RR will depend on factors such as sexual behaviour within the groups being compared and on prior knowledge about how this behaviour will affect HIV prevalence (for example, if nonrespondents 20–24 years old engaged in more risky sexual behaviour than the respondents in this age category, the RR will depend on available data from previous surveys or the published literature).

Example

To illustrate the standardization approach to adjust for nonresponse in a survey, we investigated the possible effects of the age distribution of nonrespondents, their distribution by rural and urban residence and their distribution by province using data from a hypothetical country. Table 3 shows the results. The example uses three alternative assumptions concerning the relative risk among those who were not tested compared with those who were tested: that they had the same HIV prevalence (RR = 1.0), that they had 50% lower prevalence (RR = 0.5) or that they had 50% higher prevalence (RR = 1.5).

Columns (a) and (b) show the observed distribution of people eligible for testing and those actually tested by age, residence and province. Column (c) shows the distributions of those who were eligible but who were not tested (the difference between the first two columns). Column (d) shows the HIV prevalence (%) among those who were tested, and column (e) shows the expected number of HIV-positive people among those who were tested (can be obtained by multiplying the numbers tested by the prevalence and dividing by 100). This number allows for the usual weighting of the interviewed population to represent the total national population. Column (g) shows the expected number of HIV-positive people among those not tested and is obtained by multiplying the observed prevalence (%) in column (d) by the numbers of those not tested in each category of the background characteristic. The adjusted HIV prevalence can then be estimated by adding the total number of HIV-positive people expected among those not tested to the total number of HIV-positive people among those who had been tested and dividing it by the total eligible population (such as using the first background characteristic, age, and assuming RR = 1, the total number of HIV-positive people is expected to be 214 + 569 = 783, so that the adjusted prevalence is 783/7700*100 = 10.17%). Note that this approach leads to slightly different overall adjusted prevalence depending on the background characteristic used in the adjustment (for example, assuming RR = 1, adjusted prevalence based on the age distribution, residence and province was 10.2, 10.1, and 10.4, respectively).

Column (f) shows the expected numbers of HIV-positive people among those not tested if their prevalence was half of that observed among those tested in the same age group or residence group or province. Column (h) shows the expected numbers infected among those not tested if their prevalence was 50% higher than those tested in each respective age group, residence area or province.

This procedure is essentially the same as direct standardization, where the prevalence rates by age (residence or province) in one group are applied to another. This worksheet shows all the calculations in terms of observed and expected numbers, but proportional distributions (by age, residence or province) can also be used.

Table 3. Adjustment for HIV testing nonresponse

Background characteristic	Number eligible (a)	Number tested (b)	Number not tested (c)	Observed percentage HIV positive (d)	Expected number of HIV-positive people among those tested (e)	among those not tested: assumed relative risk 0.5	1 (g)	1.5 (h)
Age (years)								
15–19	1 800	1 350	450	5.0	68	11	23	34
20–24	1 500	1 090	410	12.0	131	25	49	74
25–29	1 300	920	380	15.0	138	29	57	86
30–34	1 000	740	260	13.0	96	17	34	51
35–39	900	645	255	10.0	65	13	26	38
40–44	700	525	175	9.0	47	8	16	24
45–49	500	350	150	7.0	25	5	11	16
Total	7 700	5 620	2 080	**10.1**	569	107	214	321
Overall adjusted prevalence						**8.8**	**10.2**	**11.6**
Residence								
Urban	3 000	2 190	810	12.0	261	48	96	145
Rural	4 700	3 430	1 270	9.0	309	57	114	171
Total	7 700	5 620	2 080	**10.1**	569	105	211	316
Adjusted prevalence						**8.8**	**10.1**	**11.5**
Province								
A	1 200	936	264	0.09	84	12	24	36
B	1 500	1 213	287	0.11	133	16	32	47
C	1 800	1 295	505	0.09	117	23	45	68
D	1 200	1 056	144	0.07	78	5	11	16
E	2 000	1 120	880	0.14	157	62	123	185
Total	7 700	5 620	2 080	**10.1**	569	117	235	352
Adjusted prevalence						**8.9**	**10.4**	**12.0**

Regression analysis

A more sophisticated analysis can be performed to adjust for nonresponse using regression models. Whereas the manual method only adjusts for one background characteristic at a time, regression models can adjust for several different characteristics simultaneously.

In general, the characteristics used for the prevalence adjustment will include the variables found to significantly affect HIV prevalence in a multivariate statistical analysis using data among survey respondents.

For this adjustment, a multiple logistic regression model with HIV status (positive = 1 and negative = 0) as the dependent variable can be performed on those individuals who were interviewed and who were tested for HIV. All variables measuring relevant background characteristics expected to have an affect on HIV status can be included in the model. Typical characteristics might include age, sex, current and past urban or rural residence, region, educational level, marital status, socioeconomic status, mobility history and a variety of sexual behaviour indicators such as age when engaging in first sex, condom use and number of lifetime sexual partners. The odds ratios and *P*-values associated with each predictor variable will indicate which of the variables are significantly associated with HIV prevalence. The regression coefficients of the variables that are significantly associated with HIV prevalence can then be applied to the known characteristics of the individuals who were interviewed in the survey but not tested to predict their HIV status. Whereas the HIV status of individuals who were tested is always scored as zero or one, the predicted status for individuals who have not been tested is estimated as the probability of being infected: as a proportion between zero and one. All the scores can be averaged to produce overall estimates of HIV prevalence that combine the contributions of those tested and not tested.

For the individuals who were absent during the survey, only a few characteristics will be known that will limit the specificity of the regression analysis. The manual approach is therefore recommended to adjust for absence in a survey, whereas the regression analysis is generally recommended to adjust for test refusals.

5.6. Weighting of survey results

When data from community-based surveys are analysed, the design of the sample should be considered. Probability proportional to size sampling, where the probability of selecting a particular sampling unit is proportional to its population size, is commonly used in surveys that follow a multi-stage sampling design. Probability proportional to size is self-weighting and simplifies the calculation of prevalence: if probability proportional to size sampling methods are used, weighting is generally not required in the analysis. However, if weighting of sampling units is not considered in the design of the survey and sampling units in the population have different probabilities of being selected, the final estimates of HIV have to be weighted according to sampling design to avoid bias. Sampling weights, defined as the inverse of the probability of selection, can be used in the analysis to adjust the estimates (See section 2.4).

5.7. Estimating HIV prevalence from antenatal clinic surveillance

The primary purpose of antenatal clinic surveillance is to assess trends in the HIV epidemic over time among pregnant women. However, due to the lack of other major sources of data, antenatal clinic surveillance data are also used to estimate the level of HIV prevalence in countries.

When national estimates of HIV are calculated using data collected from antenatal clinics, the following issues need to be considered.

- How representative are women attending public antenatal clinics of all pregnant women in a country? In most countries with generalized epidemics, more than 80% of

women attend antenatal clinics (UNAIDS/WHO, 2003b). Women who do not attend antenatal clinics are often more rural, less literate and older than women attending antenatal clinics, whereas women with higher socioeconomic status may use private clinics. HIV prevalence among these women not attending public antenatal clinics is likely to be lower than among those attending (UNAIDS/WHO, 2003b).

- What is the geographical representation of the antenatal clinics included in surveillance? National surveillance systems are often based on a convenience sample of clinics, with rural areas often underrepresented. Estimates of HIV prevalence (which are likely to be higher in urban areas than in rural areas) will have to be adjusted accordingly.

5.8. UNAIDS/WHO methods to estimate national prevalence from antenatal clinic surveillance

UNAIDS and WHO, with the guidance of an external group of scientists and researchers (the UNAIDS Reference Group on Estimates, Modelling and Projections), have developed a set of methods and assumptions to model epidemic trends and to determine annual estimates of HIV prevalence in countries. For countries with generalized epidemics, that is, where HIV is firmly established in the general population, the Estimation and Projection Package (EPP) has been designed as a tool to construct national and subnational (such as urban and rural or provincial) epidemic curves, which is an essential step in estimating levels and trends in the epidemic and its impact. In these countries, the HIV prevalence among pregnant women attending antenatal clinics is taken to represent the prevalence in all adults (male and female) aged 15–49 years (Grassly et al., 2004).

For each sub-epidemic defined by the user, the EPP fits a simple epidemic model to a full set of HIV surveillance data points collected from antenatal clinics over time. This produces an estimate of the time trend of adult HIV prevalence for each sub-epidemic. The fits to individual sub-epidemics are then applied to the populations assigned by the user to each sub-epidemic to produce the prevalence estimates and trends in the overall national epidemic.

The EPP model incorporates population change over time. By varying four parameters, curves can be fitted to a variety of epidemic types, including epidemics growing slowly or rapidly, stable epidemics in which prevalence has peaked and epidemics starting to show evidence of a decline. The four model parameters – the rate of growth of the epidemic, the start year of the epidemic, the fraction of the population considered to be at risk of infection at the start of the epidemic and a behavioural response parameter that determines the final epidemic prevalence – are varied to obtain the best fit to data points.

The EPP provides the user with the ability to apply prevalence adjustments to surveillance data. For example, in countries where rural areas are underrepresented, HIV prevalence can be adjusted downward in the reported epidemic data for rural areas. Similarly, adjustment can be made to the data for national, urban and rural strata or subnational strata, using the prevalence estimates from a national survey as benchmark (discussed below). Epidemic curves produced by EPP are then used in the Spectrum software to generate estimates of national prevalence, incidence and mortality by sex and age groups (Stover, 2004a). The final estimates for the national population assume by default that the prevalence among women is 1.3 times higher than among men in generalized epidemics (Stover, 2004a).

The published literature (Ghys et al., 2004; Stover, 2004a; UNAIDS Reference Group on Estimates, Modelling and Projections, 2002; Walker et al. 2004) describes the 2003 revisions of the models in more detail. Updated models and software have been produced in March 2005 and can be downloaded from the UNAIDS (2005a) web site.

5.9. Comparing HIV prevalence from antenatal clinic surveillance with estimates obtained from population-based surveys and combined analysis to obtain a national estimate of HIV prevalence

No gold-standard method exists to measure national HIV-1 prevalence (Boerma, Ghys & Walker, 2003). Although population-based surveys with national coverage represent a much wider proportion of the population (and include both women and men) than most surveillance systems, substantial uncertainty is often associated with the HIV prevalence estimates because these surveys tend to exclude groups at increased risk of HIV and because individuals can refuse to participate or can be absent from the household at the time of the survey. The quality of the survey, including the quality of HIV testing, and the potential biases that might affect prevalence estimates need to be carefully considered and, where appropriate, HIV prevalence should be adjusted to correct for bias. Survey estimates can be adjusted for nonresponse as shown in section 5.5, but the adjusted estimates still only provide an incomplete solution because the association between absence and HIV-1 prevalence cannot be directly measured from the survey data.

Prevalence estimates obtained from antenatal clinic surveillance have shown to correlate reasonably well with prevalence of adults (age 15–49 years) in the same community (Grassly et al., 2004). However, extrapolating data collected among pregnant women to the general adult population is based on assumptions that might not apply equally to all countries. In addition, these surveys often have limited geographical coverage, especially in rural areas. The quality of antenatal clinic surveillance and hence the quality of the HIV prevalence estimates varies greatly between countries.

The value of antenatal clinic surveillance is that it provides annual HIV estimates and can therefore be used in assessing trends. Where reliable estimates of HIV prevalence from population-based surveys are available, they can be used as a means to improve antenatal clinic–based estimates of HIV prevalence and associated trends.

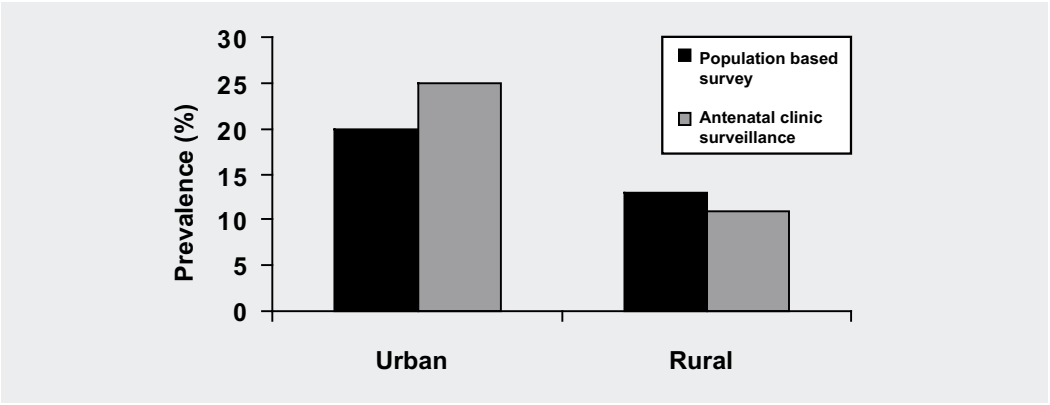
Direct comparisons between the antenatal clinic surveillance data and population-based survey results can inform the final HIV prevalence estimate in several ways.

- Comparing the HIV prevalence among pregnant women (or women who gave birth in the past two years) in the population-based survey with antenatal clinic prevalence for the same period can indicate the extent to which pregnant women in the antenatal care surveillance system are representative of all pregnant (or recently pregnant) women.
- Comparing the urban (or rural) HIV prevalence between the population-based survey and antenatal clinic data can indicate whether the urban (or rural) population was well covered by the antenatal clinic surveillance (in which case the prevalence in the adult population (15–49 years) should be similar).
- Rural (or urban) HIV prevalence can be compared between antenatal clinic survey catchment areas and nearby clusters in the population-based survey if nearby clusters can be identified.
- The ranking of regional (or geographical) areas according to HIV prevalence can be compared between antenatal clinic–based and population-based surveys. Any differences between the two may be a cause for concern. If antenatal clinic data show consistent trends in the ranking over multiple years, then the accuracy of the population-based estimates by geographical region will have to be investigated.
- Prevalence by type of antenatal care facility attended (hospital, health centre, dispensary or none) can be compared between population-based surveys and antenatal

clinic surveys. Any differences between the prevalence by type of facility could indicate differences among women in remote areas who seek care at dispensaries or who do not receive antenatal care as compared with other rural or urban women.

Fig. 3 gives an example of a comparison between HIV prevalence in an antenatal clinic survey and a demographic and health survey (both men and women combined) in the same clusters in Zambia. The prevalence estimates were fairly close in both urban and rural areas.

Fig. 3. Comparison of HIV prevalence between demographic and health survey and antenatal clinics in the same clusters in Zambia, 2001/2002



Sources: UNAIDS/WHO, 2003a.

In general, the above comparisons will show where survey-based and antenatal clinic-based HIV prevalences differ. If the differences are relatively small, no action may be necessary. However, if the differences are substantial, further investigation may be needed to understand the source of the difference and one or both estimates may need to be adjusted. For example, the rural HIV prevalence in the population-based survey may be lower than the rural prevalence from antenatal clinic surveys because more remote and less densely populated areas are often not covered by the antenatal clinic surveillance, and these areas are likely to have lower HIV prevalence. This will require adjusting the antenatal clinic prevalence. For most countries, it is recommended to adjust rural antenatal clinic HIV prevalence down by 20% (UNAIDS/WHO, 2003b).

Adjusting HIV prevalence in EPP

The UNAIDS Reference Group on Estimates, Modelling and Projections has recommended the use of HIV prevalence estimates (adjusted where necessary) from population-based surveys to calibrate the HIV prevalence levels obtained from antenatal clinic surveillance in the EPP. The latest version of the software allows this calibration to be done separately for urban and rural areas and, if data are available, by geographical regions. After the prevalence data from antenatal clinic surveillance collected over time have been entered into EPP and an epidemic curve has been fitted to the data, the curves (for all specified subpopulations) can be scaled up or down according to the HIV prevalence level in the national population-based survey. The prevalence level and year of the appropriate population-based survey can be specified for the adjustment. In addition, the prevalence in subpopulations can be scaled according to the HIV epidemiology of the population: for example, rural areas can be scaled down because of the likelihood that more remote areas are underrepresented. The EPP manual (UNAIDS, 2005b) describes the procedure in detail.

Example

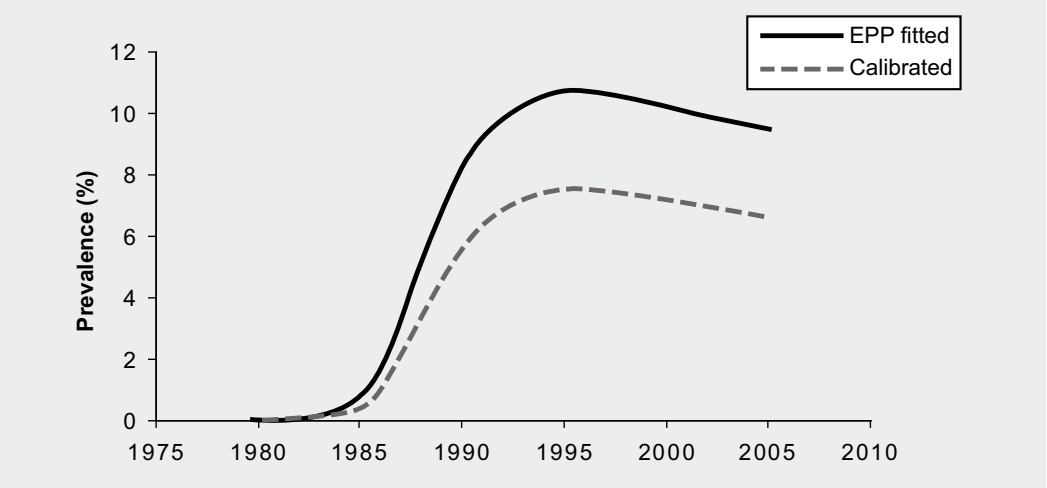
National HIV prevalence estimates for Kenya for 2003 showed differences between antenatal clinic surveillance (9.8%) and the demographic and health survey (6.7%) (Kenya Central Bureau of Statistics, Ministry of Health & ORC Macro, 2004). To reconcile these estimates, the following steps were followed (Stover, 2004b).

- The UNAIDS/WHO method (EPP) was used to fit an epidemic curve to antenatal clinic surveillance data over recent years, and the epidemic trend resulted in an estimated value for 2003 that was lower (8.5%) than the observed value (9.8%).
- Information from a recent census led to adjustment to the urban-rural population distribution. The urban weight was reduced by the new population estimates (showing a lower population in urban areas), which reduced the prevalence estimate to 8.2%.
- Based on demographic and health survey results, the assumed male-to-female ratio was corrected (the default value in Spectrum is 1.3). The Kenya demographic and health survey produced a higher ratio of 1.9 (the prevalence was 8.7% among women and 4.6% among men), which resulted in a lower estimate of male HIV prevalence and a lower overall estimate of prevalence (7.1%).
- Analysis of the Kenya demographic and health survey data showed little difference in HIV prevalence between women seeking antenatal care from hospitals, health centres and dispensaries but much lower prevalence among women who received antenatal care from another source or received no antenatal care (Stover, 2004b). The latter was more likely to occur in remotely rural areas with little access to antenatal clinics. An additional adjustment was made in the antenatal clinic–based prevalence of rural women, which resulted in an overall prevalence of 6.9%.

This example shows how different adjustments can help resolve differences between HIV prevalence estimates from population-based surveys and antenatal clinic–based data.

The revised EPP software also allows for direct adjustment of HIV prevalence estimates from population-based surveys. For the Kenya data, the epidemic curve fitted to the antenatal clinic data over recent years can be scaled down to the population-based prevalence of 6.7% in 2003, with the results shown in Fig. 4.

Fig. 4. Data obtained from antenatal clinic surveillance in Kenya scaled down to match the prevalence obtained from the demographic and health survey in 2003



6. Dissemination of results

The best way to disseminate results of a survey is through a survey report, such as demographic and health survey reports (MEASURE DHS+, 2005b).

Box 6 shows an outline for reporting on HIV prevalence. The four main areas that need to be reported on include the approach to HIV testing (including field and laboratory procedures), HIV testing coverage and nonresponse, adjusting HIV prevalence for nonresponse and HIV prevalence rates.

Box 6. Outline of the HIV prevalence section for a survey report

1. Methods of measuring HIV prevalence
   a) Ethics review
   b) Sampling methods
   c) Fieldwork procedures
   d) Laboratory procedures
   e) Linking HIV test results to individual data
2. HIV testing coverage
   a) HIV testing coverage by geographic area
   b) HIV testing coverage by background characteristics
3. Adjusting HIV prevalence estimates for nonresponse
4. HIV prevalence rates
   a) HIV prevalence by background characteristics
   b) HIV prevalence by urban or rural and region
   c) HIV prevalence by age

Methods of measuring HIV prevalence. This section must provide an overview of the ethics review, including organizational roles, how the review affected the type of testing (linked or unlinked), decisions related to the informed consent procedures and provision of voluntary counselling and testing. A description of the sampling methods should cover the stages in the sample selection, including selection of enumeration areas, household selection and selection of individuals in households. This could be brief if the report already has a section on sampling procedures in general. The description of the fieldwork should include the roles of the field team (such as who performed the HIV testing in the team), supervision of HIV testing, informed consent procedures, blood specimen collection and handling and tracking of specimens. For laboratory procedures, the testing strategy, the types of tests performed, specimen handling and quality assurance procedures should be described. Procedures for linking test results with individual information and ensuring the anonymity and confidentiality of test subjects should also be described. If the test results were unlinked, the procedures for ensuring anonymity and confidentiality of test subjects should be described.

HIV testing coverage and nonresponse. This section should describe the survey coverage of the population in terms of HIV testing. If coverage was high and nonresponse was low, little further discussion is needed. Coverage and nonresponse should be presented by urban and rural areas (Annex 6, model Table S1 shows the basic coverage table used in demographic and health survey reports), and nonresponse should also be described in terms of reasons for nonresponse, that is, absence or refusal, and refusals should be described in terms of those who were interviewed and those who were not interviewed (see Annex 6, model Table S2).

In addition to general coverage and nonresponse, coverage should also be reported in relation to background characteristics, such as coverage and nonresponse for men and women, by age group, residence, region, marital status, sexual behaviour and mobility (such as in Annex 6, model Table S3). Note that the characteristics of interest for planning and for monitoring and evaluation purposes may differ from the ones needed for research purposes and specifically for adjustment for nonresponse.

Adjusting HIV prevalence estimates for nonresponse. The report has to describe how the analysis of survey estimates dealt with nonresponse. The methods used for adjusting estimates for nonresponse and the variables used in the adjustment have to be described.

HIV prevalence rates. HIV prevalence (with 95% confidence intervals where possible) should be presented with relevant geographical breakdown (urban versus rural, and by region or province), for men and women and by age group. Presenting HIV prevalence by other background characteristics, such as marital status and sexual behaviour, will provide a broad overview of variation in HIV prevalence. Presenting HIV prevalence by smaller geographical areas will be interesting, providing that the sample sizes in these geographical areas are sufficiently large. It is also recommended to produce a graph of HIV prevalence by age for men and women to show the pattern of infection and the age at which prevalence peaks among men and women.

Communicating HIV prevalence results

Providing timely reports on HIV prevalence and related messages to stakeholders is important. If HIV prevalence estimates will be obtained from two sources (population-based surveys and sentinel surveillance), the estimates must be reconciled and adjusted before dissemination.

In countries where both sentinel surveillance and a population-based survey were conducted and where two estimates are available for the same time period, joint analysis (to produce a consistent estimate of HIV prevalence) and joint dissemination are recommended.

More details on producing a national report on HIV/AIDS can be found in the *Guidelines for effective use of data from HIV surveillance systems* (UNAIDS/WHO, 2004). When the report is complete, a national dissemination meeting should be held to inform key stakeholders, including representatives of the ministry of health, AIDS organizations, donors, other ministries working on AIDS and other interested parties, of the key results of the survey. Press conferences may be held and press releases may be distributed at that time.

References

Boerma JT, Ghys PD, Walker N (2003). Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard. *Lancet*, 362:1929–1931.

Cellule de Planification et de Statistique du Ministère de la Santé [Mali], Direction Nationale de Statistique et de l'Information & ORC Macro (2002). *Enquête démographique et de santé au Mali 2001*. Calverton, MD, ORC Macro International.

Chin J, Mann J. Global surveillance and forecasting of AIDS (1989). *Bulletin of the World Health Organization*, 67:1–11.

Colvin M, Mullick S (1997). *Draft outline of a national STD/HIV/AIDS surveillance strategy*. Durban, South Africa, Medical Research Council.

Department of Health, Education and Welfare (1979). *The Belmont report: ethical principles and guidelines for protection of human subjects research*. London, Department of Health, Education and Welfare, National Commission for Protection of Human Subjects of Biomedical and Behavioral Research.

Garcia-Calleja JM et al. (2004). A global analysis of trends in the quality of HIV sero-surveillance. *Sexually Transmitted Infections*, 80(Suppl. 1):i25–i30.

George JR et al. (1989). *Serologic assays for HIV antibody in dried-blood specimens collected on filter paper from neonates*. Atlanta, GA, United States Centers for Disease Control and Prevention.

Ghys PD et al. (2004). The UNAIDS Estimation and Projection Package: a software package to estimate and project national HIV epidemics. *Sexually Transmitted Infections*, 80(Suppl. 1): i5–i9.

Grassly NC et al. (2004). Uncertainty in estimates of HIV/AIDS: the estimation and application of plausibility bounds. *Sexually Transmitted Infections*, 80(Suppl. 1):i31–i38.

Hulley SB, Cummings SR (1988). *Designing clinical research*. Baltimore, Williams & Wilkins.

Johnson K, Way A, Otieno F (forthcoming). Nonresponse analysis of the 2003 Kenya demographic and health survey HIV prevalence data.

Kenya Central Bureau of Statistics, Kenya Ministry of Health and ORC Macro (2004). *Demographic and health survey 2003*. Calverton, MD, ORC Macro International.

Lemeshow S et al. (1990). *Adequacy of sample size in health studies*. New York, John Wiley & Sons.

Love CB, Thomson EJ, Royal CD (1999). *Ethical issues in research involving human participants*. Bethesda, MD, United States National Library of Medicine (Current Bibliographies in Medicine, 99-3; http://www.nlm.nih.gov/pubs/cbm/hum_exp.html, accessed 19 May 2005).

Lydié N et al. (2004). Mobility, sexual behaviour, and HIV infection in an urban population in Cameroon. *Journal of Acquired Immune Deficiency Syndromes*, 35:67–74.

MacQueen KM, Sugarman J (2003). *HIV Prevention Trials Network (HPTN): ethics guidance for research*. Research Triangle Park, NC, HPTN/Family Health International.

McQuillan G, Kottiri B, Kruszon-Moran D (2005). The prevalence of HIV in the United States household population: the National Health and Nutrition Examination Surveys, 1988 to 2002 (abstract). *12th Conference on Retroviruses and Opportunistic Infections, 22–25 February 2005, Boston, MA, USA* (http://www.retroconference.org/2005/CD/Abstracts/25370.htm, accessed 19 May 2005).

MEASURE DHS+ (2005a). Basic documentation [web site]. Calverton MD, ORC Macro International (http://www.measuredhs.com/methodology/basic_doc.cfm, accessed 19 May 2005).

MEASURE DHS+ (2005b). Country statistics [web site]. Calverton MD, ORC Macro International (http://www.measuredhs.com/countries/start.cfm, accessed 19 May 2005).

Office for Human Research Protections, Department of Health and Human Services (2004) [web site]. Washington, DC, Office for Human Research Protections, Department of Health and Human Services (http://www.hhs.gov/ohrp, accessed 19 May 2005).

ORC Macro (2003). *Demographic and health surveys model household questionnaire*. Version 8.9. Calverton, MD, ORC Macro International.

Respess RA, Rayfield MA, Dondero TJ (2001). Laboratory testing and rapid HIV assays: applications for HIV surveillance in hard-to-reach populations. *AIDS*, 15(Suppl. 3):S49–S59.

Sharman A (2000). *HIV testing in population-based surveys: proposed HIV field-testing protocols for MEASURE DHS+ surveys*. Calverton, MD, ORC Macro International.

Shisana O et al. (2002). *Nelson Mandela/HSRC study of HIV/AIDS: South African national HIV prevalence, behavioural risks and mass media household survey 2002*. Cape Town, Human Sciences Research Council Publishers.

Som RK (1973). *A manual of sampling techniques*. London, Heinemann Educational Books Ltd.

Stover J (2004a). Projecting the demographic consequences of adult HIV prevalence trends: the Spectrum Projection Package. *Sexually Transmitted Infections*, 80(Suppl. 1):i14–i18.

Stover J (2004b). *National HIV prevalence in Kenya*. Glastonbury, CT, Futures Group International.

Stroup DF, Teutsch SM (1998). *Statistics in public health. Qualitative approaches to public health problems*. Oxford, Oxford University Press.

Thiam M, Aliaga A (1999). A SAS macro for estimating sampling errors of means and proportions in the context of stratified multistage cluster sampling. In: *Proceedings of the Survey Research Methods Section, American Statistical Association*. Alexandria, VA, American Statistical Association (http://www.amstat.org/sections/srms/Proceedings/papers/1999_140.pdf, accessed 19 May 2005).

UNAIDS (2004a). 2004 *Report on the global AIDS epidemic*. Geneva, UNAIDS (http://www.unaids.org/bangkok2004/report.html, accessed 19 May 2005).

UNAIDS (2004b). *Responding to surveillance: methods and software to produce HIV/AIDS estimates in the era of population-based prevalence surveys*. Geneva, UNAIDS (http://www.epidem.org/Publications/Glion%20report%20final.pdf, accessed 19 May 2005).

UNAIDS (2005a). Methods and assumptions for estimates [web site]. Geneva, UNAIDS (http://www.unaids.org/Unaids/EN/Resources/Epidemiology/How+do+UNAIDS_WHO+arrive+at+estimates/Epi_Methods.asp, accessed 19 May 2005).

UNAIDS (2005b). Epidemiological software and tools [web site]. Geneva, UNAIDS (http://www.unaids.org/Unaids/EN/Resources/Epidemiology/EPI_SoftwareTools.asp, accessed 19 May 2005).

UNAIDS/WHO (1999). *Operational characteristics of commercially available assays to determine antibodies to HIV-1 and/or HIV-2 in human sera*. *Report 11*. Geneva, WHO/UNAIDS (http://www.who.int/diagnostics_laboratory/publications/evaluations/en/, accessed 19 May 2005).

UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2000). *Guidelines for second generation HIV surveillance*. Geneva, UNAIDS/WHO (http://www.unaids.org/EN/resources/epidemiology/epi_recent_publications/secondgensurveillance2000.asp, accessed 19 May 2005).

UNAIDS/WHO Working Group on Surveillance (2001). *Guidelines for using HIV testing technologies in surveillance*. Geneva, UNAIDS/WHO (http://www.who.int/hiv/pub/epidemiology/pub4/en, accessed 19 May 2005).

UNAIDS/WHO (2002a). *HIV simple/rapid assays: operational characteristics: whole blood specimens*. *Report 12*. Geneva, WHO/UNAIDS (http://www.who.int/diagnostics_laboratory/publications/evaluations/en, accessed 19 May 2005).

UNAIDS/WHO (2002b). *HIV assays: operational characteristics: urine specimens, oral fluid (saliva) specimens*. *Report 13*. Geneva, WHO/UNAIDS (http://www.who.int/diagnostics_laboratory/publications/evaluations/en, accessed 19 May 2005).

UNAIDS/WHO (2003a). *Reconciling antenatal clinic–based surveillance and population-based survey estimates of HIV prevalence in sub-Saharan Africa*. Geneva, UNAIDS/WHO (http://www.who.int/hiv/strategic/mt170203/en, accessed 19 May 2005).

UNAIDS/WHO (2003b). *Guidelines for conducting HIV sentinel serosurveys among pregnant women and other groups*. Geneva, WHO/UNAIDS (http://www.who.int/hiv/pub/surveillance/anc_guidelines/en, accessed 19 May 2005).

UNAIDS/WHO (2004). *Guidelines for effective use of data from HIV surveillance systems*. Geneva, UNAIDS/WHO (http://www.who.int/hiv/strategic/surveillance/hivpubepidemiology/en, accessed 19 May 2005).

UNAIDS Reference Group on Estimates, Modelling and Projections (2002). Improved methods and assumptions for estimation of the HIV/AIDS epidemic and its impact: recommendations of the UNAIDS Reference Group on Estimates, Modelling and Projections. *AIDS*, 16:W1–W14.

UNICEF (2000). *Monitoring progress toward the goals of the World Summit for Children: end-decade multiple indicator survey manual*. New York, UNICEF (http://www.childinfo.org/MICS2/finman/M2finM.htm, accessed 19 May 2005).

United States Centers for Disease Control and Prevention (1988). Perspectives in disease prevention and health promotion update: universal precautions for prevention of transmission of human immunodeficiency virus, hepatitis B virus, and other bloodborne pathogens in health-care settings. *MMWR Morbidity and Mortality Weekly Report*, 37(24):377–388.

United States Centers for Disease Control and Prevention (1999). Guidelines for national human immunodeficiency virus case surveillance, including monitoring for human immunodeficiency virus infections and acquired immunodeficiency syndrome. *MMWR Recommendations and Reports*, 48(RR-13):1–27, 29–31.

Urassa M et al. (2004). Integrated ante-natal HIV and behaviour surveillance in northern Tanzania. *XV International AIDS Conference, Bangkok, Thailand, 11–16 July 2004* (abstract MoPeC3577; http://www.aegis.com/conferences/15wac/MoPeC3577.html, accessed 19 May 2005).

Walker N et al. (2004). Estimating the global burden of HIV/AIDS: what do we really know about the HIV pandemic? *Lancet*, 363:2180–2185.

WHO (2003). *Laboratory biosafety manual*. 2nd ed., revised. Interim guidelines. Geneva, World Health Organization (http://www.who.int/csr/resources/publications/biosafety/who_cds_csr_lyo_20034/en, accessed 19 May 2005).

WHO (2005a). HIV test kit evaluations [web site]. Geneva, World Health Organization (http://www.who.int/diagnostics_laboratory/evaluations/hiv/en, accessed 19 May 2005).

WHO (2005b). Communicable Disease Surveillance and Response [web site]. Geneva, World Health Organization (http://www.who.int/csr/en, accessed 19 May 2005).

WHO/UNAIDS (2004a). *Ethical issues to be considered in second generation surveillance*. Geneva, World Health Organization (http://www.who.int/hiv/pub/epidemiology/en/sgs_ethical.pdf, accessed 19 May 2005).

WHO/UNAIDS (2004b). *HIV assays: operational characteristics. Antigen/antibody ELISAs.* Geneva, World Health Organization (http://www.who.int/diagnostics_laboratory/publications/evaluations/en, accessed 19 May 2005).

Yoder PS, Konaté MK (2002). *Obtaining informed consent for HIV testing: the DHS experience in Mali*. Calverton, MD, ORC Macro International.

Zaba B et al. (2004). How well do cross-sectional population surveys measure HIV prevalence? Exploring the effects of non-participation. *XV International AIDS Conference, Bangkok, Thailand, 11–16 July 2004* (abstract LbOrC23; http://www.aegis.com/conferences/15wac/LbOrC23.html, accessed 19 May 2005).

Zuma K et al. (2003). Risk factors for HIV infection among women in Carletonville, South Africa: migration, demography and sexually transmitted diseases. *International Journal of STD and AIDS*, 14:814–817.

Annex 1. Example of a survey budget

Description	Quantity	Units	Unit cost	Total cost
1. Survey organization				
Salaries				
Survey director				
Survey specialists				
Accountant				
Administrative assistant				
Driver				
Compensation for daily expenses				
Survey director				
Survey specialists				
Driver				
Other				
Communication				
Office supplies				
2. Household listing				
Salaries				
Cartographers				
Drivers				
Compensation for daily expenses				
Cartographers				
Drivers				
Training				
Room rental				
Supplies				
2. Pretest and training				
Salaries				
Trainers				
Team leaders				
Interviewers and editors				
Other				
Room rental				
Pretest and training questionnaires (photocopies)				
Supplies				

Description	Quantity	Units	Unit cost	Total cost
3. Fieldwork				
Salaries				
Team leaders				
Interviewers and editors				
Drivers				
Compensation for daily expenses				
Team leaders				
Interviewers and editors				
Drivers				
Biological sample collection and shipment				
Disposable gloves				
Lancets				
Sample containers (such as dried blood spot cards or serum, saliva or urine containers)				
Bar code reader and software				
Biohazard containers and bags				
Other HIV testing supplies				
Other				
Fieldwork supplies				
4. Data entry				
Salaries				
Supervisor				
Office editors				
Data entry operators				
Other				
Supplies				
5. Transport				
Fuel				
Supervision				
Household listing				
Pretesting				
Fieldwork				
Other				
Vehicle maintenance				
Vehicle rental				

Description	Quantity	Units	Unit cost	Total cost
6. Print materials				
Publicity				
Survey manuals				
Household questionnaire				
Women's questionnaire				
Men's questionnaire				
7. HIV tests				
Salaries				
Assay costs				
Other laboratory supplies				
8. Mobile voluntary counselling and testing (if provided)				
Salaries				
Vehicle costs (rental, fuel and maintenance)				
Assay costs				
Other laboratory supplies				
9. Other expenses				
Consultants				
Fieldwork insurance				
Miscellaneous				
Total				
Overhead				
Fees				
Grand total				

Annex 2. Informed consent forms

a) Example of obtaining informed consent for the interview

Instruction to the interviewer: read this consent form aloud and clearly to the respondent.

Good morning/good afternoon. my name is [NAME OF THE INTERVIEWER]. I am representing [FOR EXAMPLE, THE MINISTRY OF HEALTH].

The Ministry of Health and their partners are working together on preventing and controlling HIV and AIDS. The Ministry of Health would like to know what is happening in the general population in order to be able to improve and plan the delivery of services to our people. The Ministry is getting this information by asking people in their homes some questions. I am here to ask you some questions. Some of these questions will be about your personal life. I am aware that some of the questions are sensitive, but all the information you give me will be kept strictly confidential. Participation in this survey is voluntary. You can refuse to answer all or some of the questions, but the Ministry of Health would appreciate your help in answering all the questions. It is important for you to know that your participation will not affect your ability to use health facilities or any other services. We are hoping that you will participate since your participation and views are highly valued and important.

At this time, do you want to ask me anything about the survey? May I begin asking you the questions now?

[Instruction to the interviewer: wait for the answer and make sure you do not rush the respondent to answer. If the respondent is not answering, gently ask the question again until you get an answer.]

I have been informed about this survey and understand its purpose and objectives. I understand the details, have been informed about the requirements and hereby agree to participate in the survey.

Signature of respondent _____ Date _____

Signature of interviewer_____ Date _____

b) Example of informed consent for HIV testing

Instruction to the interviewer: the person taking the specimen should seek this consent after the interview. If this is a different person from the interviewer, he or she should start also by greeting and introducing himself or herself to the respondent. If it is the same person, he or she should go straight to the consent form and read it aloud and clearly to the respondent.

For children (younger than 18 years), the parent or the guardian must first be asked for consent. Only if the parent or guardian agrees will the child be asked to consent to the test.

As part of this survey, we want to know how many people have the HIV virus that causes AIDS.

We are asking people all over the country to give a few drops of blood for an HIV test.

We will do this in a completely safe way. All the materials we use are new, sterile and clean.

We are not putting your name on the specimen or on this paper [the questionnaire]. This means that we cannot identify the person whose blood is drawn and we cannot give people the results of the test. Information from this survey will be made public, but no names or identification of any person will be used anywhere in the report or otherwise made public.

Interviewer: offer voluntary counselling and testing referral card as you explain about voluntary counselling and testing.

If you wish to know your HIV status, I am giving you a card to go to the nearest voluntary counselling and testing centre, where you can receive free counselling and your blood can be drawn for a free HIV test. At these centres you can meet trained staff that will be available to discuss with you all issues and matters regarding HIV and AIDS.

You can refuse to have your blood drawn. The choice is yours, but the Ministry of Health would appreciate your participation in this survey. Even if you do not agree to have a specimen taken, you are still free to use the voluntary counselling and testing or any other services.

At this time, do you want to ask me anything about the survey? Do you agree to be tested?

Instruction to the interviewer: wait for the answer and make sure you do not rush the respondent to answer. If the respondent is not answering, gently ask the question again until you get an answer.

I have been informed about this survey and understand its purpose and objectives. I understand the details, have been informed about the requirements and hereby agree to have my blood drawn.

Signature of respondent _____ Date _____

Signature of person drawing blood _____ Date _____

(For children (younger than 18 years), the parent or guardian should also sign)

Signature of parent or guardian_____ Date _____

Annex 3. Example of survey staff and job descriptions

Staff type	Job responsibilities
Survey management and technical personnel	
Survey director	Overall survey and financial management, technical lead
Survey specialists and consultants	Survey design, sampling methods, training, field supervision and quality control, data management, statistical analysis and report writing
Accountant	Maintains all survey accounts, handles accounts receivable and payable, monitors spending, monitors vehicle use and schedules maintenance
Support staff	
Administrative assistants	Provide general administrative support to survey staff
Drivers	Driving and track fuel and mileage
Household listing	
Cartographers and listing teams	Find enumeration area, map buildings in enumeration area and provide sampling frame
Fieldwork	
Team leaders	General supervision of team, communicate with community leaders, manage interviewer assignments, field logistics and quality control
Field editors	Keep track of questionnaires and edit questionnaires
Interviewers (male and female)	Interview and obtain specimens for HIV testing
HIV test technicians	Obtain specimens for HIV testing in cases where interviewers are not allowed to take blood specimens
Data entry and processing	
Data manager	Manages data entry programme development, data entry, data editing and data cleaning
Office editor	Logs in questionnaires, performs final hand edits and coding of questionnaires
Data encoders	Enter data from questionnaires into electronic data files
Statistician	Performs statistical analysis of data

Annex 4. Using dried blood spots

Storage of dried blood spot samples is undesirable in very humid (>50% relative humidity) and hot conditions, although there should be little loss of reactivity if samples are stored for less than 14 days in high humidity (Sharman, 2000). Samples can be stored in their bags for up to 30 days at room temperature and for up to 90 days at 4°C (George et al., 1989). If they are to be kept for more than 90 days, bags have to be stored at –20°C.

Dried blood spot collection and storage procedure

1. Set up the sample collection site. Blood should preferably be collected indoors. Make sure there is adequate light. A couch, bed or mat should be easily accessible if the respondent feels faint during the procedure and needs to lie down.
2. Seat the subject and record information pertinent to study on a specimen tracking form.
3. Read the informed consent statement and ask the participant to sign it. The technician can sign it if the participant is illiterate and oral consent is provided. Place the first bar code sticker next to the informed consent signature in the household questionnaire.
4. Put on a pair of disposable gloves.
5. Carefully remove a new filter paper card from a package, making sure not to touch the areas in the preprinted circles.
6. Once a package of filter paper cards has been opened, it should be stored in a zipper lock bag to prevent exposure to moisture or other environmental factors.
7. Place the card on a dry surface.
8. Hold the participant's hand palm side up below his or her heart to increase blood pressure to the hand. Massage the finger (middle or ring finger) to cause blood to accumulate in the tip of the finger.
9. Cleanse the finger pad (not just the tip or side) with an alcohol swab. Allow the finger to completely dry.
10. Firmly prick the finger pad with a sterile lancet.
11. Place the used lancet in the biohazard container.
12. Wipe off the first drop of blood with a sterile gauze pad. This first drop contains tissue fluids that dilute the sample.
13. Gently touch the next large drop of blood with the filter paper card. Make sure the subject's finger does not touch the card. Apply gentle pressure, but do not squeeze the finger to increase blood flow. Allow the blood to soak the paper and fill the preprinted circle. Avoid touching or smearing the spots.
14. Use only one application of blood per circle. Do not layer the sample in an attempt to fill the circle. If the circle has not been adequately filled and flood flow has diminished, repeat steps 8–13 on another finger using the next free circle on the card.
15. Collect enough blood to spot at least three circles on the card. Two to three drops of blood should be sufficient for HIV testing.
16. Place a plaster over the pierced finger.
17. Attach the second bar code sticker on the back of card in the space between circles.
18. Air-dry the specimens for four hours at room temperature. During drying, avoid stacking the cards or placing them on absorbent surfaces. Place the cards in portable collection boxes to hold the cards horizontally during drying and protect the cards from exposure to sunlight and other environmental factors. Place pouches of regeneratable desiccant at the bottom of the boxes to assist with the drying process. The boxes should only be used for drying and not for long-term storage.
19. When the specimens are completely dry, stack them in zipper-lock freezer bags with glassine paper between each card. Place a nonregeneratable desiccant pouch and humidity card in the bag.

Annex 5. Handling and disposing of biohazardous waste

Proper handling and disposal of biohazardous waste produced by HIV testing is essential to ensure that individuals do not come into contact with this waste in a way that could expose them to HIV or other pathogens. Biohazardous waste includes any materials (lancets, gloves, gauze, swabs, etc.) that may be contaminated with any body fluids such as blood or sera (Sharman, 2000).

Laboratories performing HIV testing should already have procedures in place for handling biohazardous waste. Basic biohazardous waste handling involves storage of waste in clearly marked biohazard containers until this waste is disposed of. Typically there will be biohazard bins with biohazard bags that may be removed for disposal. These bins should be periodically disinfected. Sharp items should be stored in puncture-resistant containers. Disposal of biohazardous waste related to HIV testing involves autoclaving or incinerating the biohazard bags or sharps containers along with the biohazardous contents.

In the field, handling and disposal of biohazardous waste becomes more problematic (Sharman, 2000). Sharp items will be stored in puncture-resistant containers lined with biohazard bags. Other materials contaminated with blood will be stored in clearly marked containers lined with biohazard bags. If a health facility or commercial centre for biohazardous waste disposal is available, the waste will be taken to that facility. Otherwise the field team must remove the biohazard bags and incinerate the bags along with the contents. Precautions must be taken not to be exposed to the biohazardous materials during this process. Biohazardous waste must be properly disposed of at the end of each day and must not be transported from one survey site to another.

Annex 6. Reporting on coverage of interviews and HIV testing

Model tables on how to present key information related to nonresponse from surveys that include HIV testing.

Model Table S1. Results of household and individual interviews			
Number of households, number of interviews and response rates, according to residence			
	Residence		
	Urban	**Rural**	**Total**
Household interviews			
Households sampled			
Households occupied			
Households interviewed			
Household response rate			
Individual interviews: women			
Number of eligible women			
Number of eligible women interviewed			
Response rate for eligible women			
Individual interviews: men			
Number of eligible men			
Number of eligible men interviewed			
Response rate for eligible men			

Model Table S2. HIV testing coverage			
Coverage of HIV testing for men and women by residence			
	Residence		
	Urban	**Rural**	**Total**
Women			
Tested			
Not tested			
Absent			
Refused			
Interviewed			
Not interviewed			
Missing or other			
Total			
Men			
Tested			
Not tested			
Absent			
Refused			
Interviewed			
Not interviewed			
Missing or other			
Total			
Both sexes			
Tested			
Not tested			
Absent			
Refused			
Interviewed			
Not interviewed			
Missing or other			
Total			

Model Table S3. Women's and men's HIV testing coverage by background characteristics					
Coverage of HIV testing by background characteristics					
	Tested	**Refused**	**Absent**	**Missing or other**	**Total**
Age (years)					
15–19					
20–24					
25–29					
30–34					
35-39					
40–44					
45–49					
50–54					
55–59					
Residence					
Urban					
Rural					
Region					
North					
South					
East					
West					
Marital status					
Single					
Married (first time)					
Remarried					
Widowed, separated or divorced (not remarried)					
Higher-risk sex in past year					
Yes					
No					
Condom use during last higher-risk sex					

Yes	
No	
Commercial sex in past year	
Yes	
No	
Condom use during last commercial sex	
Yes	
No	
Time away from household in last 12 months None ≤1 month >1 month	

For further information, please contact:
World Health Organization
Department of HIV/AIDS
20, avenue Appia CH-1211 Geneva 27 Switzerland
E-mail: hiv-aids@who.int
http://www.who.int/3by5

ISBN 92 4 159370 9