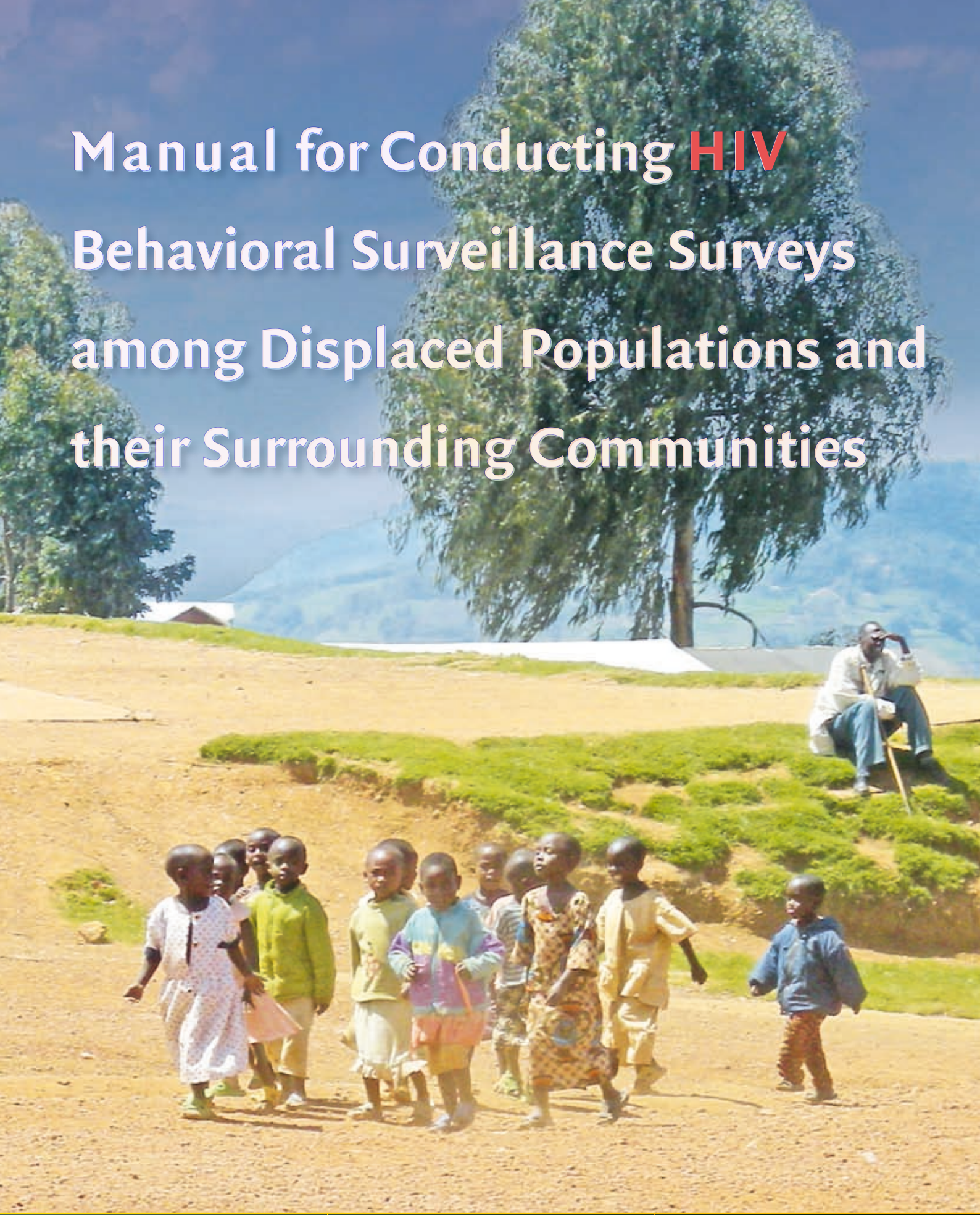
Manual for Conducting HIV Behavioral Surveillance Surveys among Displaced Populations and their Surrounding Communities



UNHCR
The UN Refugee Agency

Great Lakes Initiative on AIDS (GLIA)

THE WORLD BANK

Contact details for further information:
- Public Health and HIV Section UNHCR, hivaids@unhcr.org
- Global AIDS Monitoring and Evaluation Team (GAMET), The World Bank gamet@worldbank.org
- GLIA Secretariat, info@greatlakesinitiative.org

Cover and interior page design by Frans Aussems (Haren, The Netherlands)

Cover photo by Marian Schilperoord

United Nations High Commissioner for Refugees

Great Lakes Initiative against HIV/AIDS

World Bank

Manual for Conducting HIV Behavioral Surveillance Surveys among Displaced Populations and their Surrounding Communities

March 2008

UNHCR
The UN
Refugee Agency

THE WORLD BANK

Table of Contents

Acknowledgements

This manual is the result of the collective effort of public health organizations, agencies and individuals concerned with strengthening the response to HIV and AIDS among populations affected by conflict. Particular thanks are due to Paul Spiegel, Chief, Public Health and HIV Section at UNHCR Geneva who conceived of and guided each step of the manual and Sara Hersey, UNHCR/GLIA consultant, who wrote and edited the manual's contents and provided technical assistance in survey design, implementation and analysis.

Marian Schilperoord and Njogu Patterson of UNHCR, Marelize Görgens-Albino of the World Bank, Reinhard Kaiser of the European Centres for Disease Control, and Sophia Luhindi of GLIA provided valuable advice during the development of this manual. Elizabeth Rowley of the Johns Hopkins School of Public Health generously shared her experience in Tanzania.

Kathleen McDavid of the United States Centers for Disease Control assisted greatly in refining and testing the survey instruments. The survey teams in Uganda provided invaluable assistance in detailing lessons learned from the field, including Johanna Claass, UNHCR Geneva, Judith Bamuturake and Tina Ghelli, UNHCR Kampala, Michael Muyonga, Ugandan Ministry of Health, the UNHCR Mbarara and Hoima field offices, and the site supervisors and interviewers in Nakivale, Oruchinga, and Kyangwali.

Thanks also to Richard Seifman, Curtis Blanton, Dieudonne Yiweza, Laurie Bruns, Rick Brennan, Cheikh Tidiane Toure, and Francesco Checchi for their technical input into this work.

List of acronyms

AIDS	Acquired immune deficiency syndrome
BSS	Behavioral surveillance surveys
DHS	Demographic and Health Survey
EPI	Expanded Programme on Immunization
GLIA	Great Lakes Initiative on HIV/AIDS
HIV	Human immuno-deficiency virus
MGD	Millennium Development Goals
MOH	Ministry of Health
NGO	Non-governmental organization
PEPFAR	President's Emergency Plan for AIDS Relief
PLHIV	People living with HIV/AIDS
PMTCT	Prevention of mother-to-child transmission
PPS	Probability proportionate to size
PSU	Primary sampling unit
STI	Sexually transmitted infection
UN	United Nations
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNGASS	United Nations General Assembly Special Session on HIV/AIDS
UNHCR	United Nations High Commissioner for Refugees
VCCT	Voluntary and confidential counseling and testing
WHO	World Health Organization

Chapter 1

Introduction

While the past decade has seen great strides in expanding HIV and AIDS prevention, care and treatment programs across Africa and the rest of the world, populations displaced by conflict have been noticeably absent from many of the country and regional-level initiatives designed to combat the disease. The Great Lakes Initiative on HIV/AIDS (GLIA) was created to foster regional cooperation, complementing the efforts of national AIDS commissions of the 6 GLIA countries. It has the following objectives:

- Strengthen collaboration and cooperation among the GLIA countries (Burundi, Democratic Republic of Congo, Rwanda, Uganda, and Tanzania) and agencies in the Great Lakes region for a strong and coordinated response;
- Promote dialogue and sharing among GLIA countries;
- Support innovative and catalytic projects responding to the priority needs of GLIA;
- Encourage all partners, particularly PLHIV, the private sector, NGOs and communities, UN agencies, and donors to participate in an expanded and coordinated response in the Great Lakes region;
- Strengthen regional coordination to ensure better use of resources and to mobilize additional financial resources; and
- Establish a regional forum to share resources and best practices.

In order to strengthen evidence-based program and policy responses that support these objectives, GLIA instituted activities to gather standardized and high quality data on HIV prevalence and behavioral trends among populations that it targets. Among these populations are displaced or mobile populations.

Routine surveillance of HIV prevalence has been recognized as a necessary tool in monitoring and mitigating the effect of the disease ever since the early stages of the AIDS epidemic. Two decades later, it remains of utmost importance to know how many people have acquired HIV and where they are located. However HIV sero-surveillance does not provide adequate information on what risk factors may be driving infections and what interventions would be most effective in stopping this transmission.

The Behavioral Surveillance Survey (BSS) is a globally recognized tool used to measure factors that may lead to HIV exposure and to evaluate programmatic responses. However, to date few quality BSS have been conducted among populations displaced by conflict and their surrounding host communities, and no templates existed to guide behavioral data collection in these unique populations. This manual is the result of the work of GLIA, the United Nations High Commissioner for Refugees (UNHCR) and the World Bank in designing standardized BSS tools for displaced populations and providing guidelines for survey implementation – for use by the GLIA and other institutions that may need to know more about HIV-related behaviour, knowledge and attitudes amongst displaced populations.

Recognizing that a combination of methods – measuring HIV prevalence as well as measuring HIV behavioral risk factors – would provide greater insight into the dynamics driving HIV transmission and the most effective means to combat even more widespread transmission, the Joint United National Programme on HIV/AIDS (UNAIDS) and the World Health Organization (WHO) recommended that HIV surveillance should be coupled with additional surveillance of behavioral risk factors.[1]

Behavioral Surveillance Surveys (BSS) are now a recognized tool for measuring a number of factors that may contribute to HIV transmission including the following:

- Behavioral risk factors, such as multiple sex partners, unprotected sex, and injection drug use that may lead to HIV transmission
- HIV co-factors including prevalence of circumcision and symptoms of sexually transmitted infections
- Interactions between populations with high HIV prevalence and those with lower HIV prevalence
- Access to condoms and acceptability of condom use
- Access to and utilization of Voluntary and Confidential Counseling and Testing (VCCT) services
- Exposure to specific HIV interventions such as peer education or a Prevention of Mother-to-Child Transmission (PMTCT) program

Although the modes of HIV transmission are well-known, the prevalence of and factors affecting risky sexual and drug-taking behavior vary significantly among populations and areas of the world. Appropriate intervention strategies require that specific, local data are available in order to inform programmatic, resource and policy decisions and to maximize their effect.

While many countries have established routine behavioral surveillance within their national HIV monitoring and evaluation system, refugees and other displaced populations often fall outside of the purview of the host country governments. In addition, because they have been exposed to conflict situations, forced migration from their home, and often years – or decades – of displacement from their communities, displaced populations may have become vulnerable to HIV exposure in ways that behavioral surveys designed for members of a country's general population may not pick up.

Other economic, social and health needs may preclude the rapid inclusion of HIV interventions in the services provided to refugees and other displaced populations. However, national and international organizations and local governments have increasingly recognized that HIV programs are critical to the health and well-being of refugees and other displaced populations as well as the surrounding host communities with whom they live. Information on behavioral risk factors specific to these populations is of utmost importance, but

1 *Second generation surveillance for HIV: The next decade*, UNAIDS/WHO Working Group on Global HIV/AIDS/STI Surveillance, 2000, http://www.who.int/hiv/pub/surveillance/pub3/en/index.html

standards for conducting high quality BSS among displaced populations were not previously available.

This manual is the culmination of work done by a number of leading international experts in standardizing the tools and methods used for collecting behavioral data among refugees, other displaced populations, and their surrounding communities. The manual has also benefited from extensive piloting in two separate sites. It provides detailed information on how to conduct BSS in these populations and includes the following information:

- Objectives and core indicators for BSS among refugees, displaced populations and their surrounding communities
- Designing and preparing for a BSS
- Sample size calculations and sampling methods
- Standardized questionnaires and informed consent and methods for refining, translating and testing the tools
- Data collection
- Data management and analysis
- Reporting and disseminating BSS data
- Research ethics

While this manual has been designed to help standardize the survey process and describe methods for increasing BSS data quality, it is also recognized that all survey populations and areas, and site-specific realities that govern them, are different. This manual should be used as an important instrument in designing and implementing your surveys, but it should not replace the all-important local knowledge and common sense that you will bring to your own surveys.

Chapter 2

Objectives of the BSS and core indicators

Objectives of the BSS

Behavioral Surveillance Surveys can have multiple objectives. Some objectives, such as describing HIV behavioral risk factors in a population, will be common to all surveys. However, your survey may also have other purposes unique to your own situation, such as evaluating specific program components, which will you will also need to consider. Your objectives will drive the survey design, including the sample size, as well as what information will be collected by the questionnaire, and it is important to carefully consider and document them *before* moving onto the design stage of your survey.

For the purpose of this manual, the key objectives of the BSS among displaced populations and their surrounding communities are the following:

Objective 1: Establish baseline behavioral, knowledge and intervention exposure data in the displaced and surrounding populations.

The primary objective of the BSS is to describe sexual and drug using behaviors, sexual and gender-based violence, knowledge and attitudes about HIV, and access and exposure to interventions and other services in vulnerable and most-at-risk populations. This is particularly important for the GLIA, as it has a set of indicators against which to measure progress with its activities on a regular basis.

Objective 2: Estimate the type and frequency of interactions between the displaced and surrounding populations.

The BSS also measures how often and in what situations displaced populations and their host communities interact. It is particularly important to measure these interactions when HIV prevalence is higher in one of the survey populations because this data will provide information about the potential for HIV to 'cross over' to the neighboring population.

Objective 3: Provide data to allow comparison of indicators with other areas of the country and region.

Although each BSS survey will need to be designed to fit the local context, this manual presents standardized methods and tools for measuring core indicators of HIV behavioral risk. If the fundamentals of the BSS survey, as laid out in this manual, are followed, then it should be possible to compare the results of your survey to other populations and areas of the country and region.

Objective 4: Improve the understanding of HIV risk behaviors and vulnerability before and during displacement among the displaced population.

While the BSS is not able to measure risk behaviors that occurred well in the past with a high level of accuracy, it is able to describe what factors may have put the displaced population at risk for HIV transmission prior to displacement and during the displacement phase.

Objective 5: Measure trends over time.

Another key objective of the BSS is to measure changes in behavior, knowledge, attitudes and access to services over a period of time. These data will assist in monitoring and evaluating program achievements and strengthening interventions. The BSS surveys in this manual are designed to collect baseline data and then to be replicated after a period of time, usually 2-3 years, to measure trends.

Objective 6: Provide data for use by program managers and policy makers.

The BSS provides the evidence necessary for program managers and policy makers to make effective decisions regarding intervention needs and resource allocation and to leverage support and funding to fill the gaps.

Other objectives of your BSS may include providing baseline and follow-up data to evaluate specific interventions such as a Voluntary and Confidential Counseling and Testing (VCCT) service that has been scaled-up or a Prevention of Mother-to-Child Transmission (PMTCT) program that has been introduced in the area. Because sampling for BSS surveys is designed to be representative of the entire displaced and surrounding community population, they can also provide you with data on programmatic coverage in the survey catchment area.

Core indicators

The BSS for displaced populations and surrounding communities is designed to collect data on many different areas to inform HIV prevention and care programs. Broadly, this information includes the following:

- Descriptive information including age, nationality, education, religion, marital status, military activity and employment
- Population displacement and current population movement, as well as interactions between the displaced and host communities
- Alcohol and drug use
- Sexual history, recent sexual partnerships, and condom use
- Gender and sexually-based violence
- Co-factors for HIV transmission including circumcision and symptoms of sexually transmitted infections
- Knowledge, opinions and attitude towards HIV
- Exposure and access to programs and services

All of the data collected in the survey is important, but only some of the variables will measure the core indicators of risk that are of greatest interest to your program. Which indicators you chose will depend on two main factors. Firstly, what are the objectives of your survey? If the survey will provide data to evaluate a specific project, the indicators should reflect the core components of the program. For example, if your intervention efforts are focused on decreasing the number of sex partners among the youth population, an important BSS indicator will be the proportion of the population between the ages of 15 and 24 who report multiple sex partners. Another indicator of interest may be the proportion of the youth population who report any sex partner other than a spouse.

The second main factor that will influence the choice of indicators is what risk factors are driving HIV transmission in your survey area. Ideally, these risk factors will already be addressed in your program design and survey objectives, even if a lot is not yet known about them. For example, if drug use, particularly injection drug use, or transactional sex is common in the survey populations, you should select indicators that accurately measure these behaviors. At least a basic knowledge of your population and their socio-sexual culture is necessary in order to select the most appropriate BSS indicators.

A great deal of work has been done in defining and standardizing HIV behavioral risk indicators, and various international agencies and working groups have produced guidelines for constructing and measuring core HIV indicators. For more information on international standards for HIV indicators, refer to the following documents and databases:

HIV Surveys Indicators Database, MEASURE Evaluation
http://www.measuredhs.com/hivdata/

Guidelines for the Construction of Core Indicators, United Nations General Assembly Special Session on HIV/AIDS (UNGASS), UNAIDS, 2008
http://data.unaids.org/pub/Manual/2007/20070411_ungass_core_indicators_manual_en.pdf

A Framework for Monitoring and Evaluating HIV Prevention Programmes for Most-At-Risk Populations, UNAIDS, 2006
http://data.unaids.org/pub/Manual/2007/20070420_me_of_prevention_in_most_at_risk_populations_en.pdf

Millennium Development Goals Indicators, United Nations Statistics Division, 2003
http://millenniumindicators.un.org/unsd/mdg/Host.aspx?Content=Indicators/OfficialList.htm

Monitoring and Evaluation Toolkit: HIV/AIDS, Tuberculosis and Malaria, The Global Fund to Fight AIDS, Tuberculosis and Malaria, January 2006
http://www.theglobalfund.org/en/links_resources/library/evaluation_framework/

In addition to considering the internationally standardized indicators, it is also necessary to take into account indicators that are being measured by the Ministry of Health monitoring and evaluation system, indicators that are a priority for project funders, and – most importantly – indicators that will best measure the progress of interventions in your area. Ideally,

international, national, development partner and program-level indicators should be the same or at least reflect the same guiding principles. Realistically, while major advances have been made in recent years to simplify and standardize HIV risk measurements, the indicators that you ultimately select must be appropriate for your population and the epidemic dynamics in your area and for your program.

Once you have selected the important outcome measures for your local situation, you must clearly define the indicators and how they will be calculated. Then you need to ensure that your questionnaire is collecting the correct information for measuring them. In 2005 and 2006, a group of international epidemiologists and program managers reviewed the global HIV risk indicators referred to above and considered what indicators would be most appropriate as measures of HIV risk in displaced populations and their surrounding communities. The group selected a number of globally recognized indicators measuring sexual behavior, HIV testing, STI health facility utilization, and knowledge, attitudes and misconceptions as the standards for BSS in these populations. In addition, they defined several important indicators specific to displacement situations.

In Table 1 below, the core indicators for BSS among displaced populations and their surrounding communities are provided including their definition, construction, and whether they are internationally reported indicators. It is important to note that knowledge about HIV and measuring HIV risk is continually evolving. International reporting standards may change, and national government and donor requirements may be different from what is presented below. In addition, your survey population may have other important indicators, such as those specific to drug injection, which will be necessary to measure. The list of indicators is provided as guidance and should not be used verbatim without first considering your local situation. More information on analyzing each indicator is provided in *Chapter 11: Data Analysis*.

Table 1:

Indicator	Definition	Construction	International reporting
Sexual behavior			
1. Young men and women aged 15-24 who have had sexual intercourse before the age of 15	Percent of men and women aged 15-24 who had sex before the age of 15	Reported age at first sex <15 years old Denominator: Population aged 15-24, disaggregated by sex and age groups 15-19, 20-24, 15-24	UNGASS

Indicator	Definition	Construction	International reporting
2. Never-married young people aged 15-24 who have never had sex	Percent of men and women aged 15-24 who have never been married and never had sex	Reported never had sex Denominator: Population aged 15-24 who has never been married, disaggregated by sex and age groups 15-19, 20-24, 15-24	PEPFAR GLIA
3. More than one sex partner in the past 12 months among men and women aged 15-49	Percent of men and women aged 15-49 who report having sex with more than one regular, non-regular and/or transaction partners	Composite indicator of people reporting two or more sex partners in the past 12 months: (calculated through reported number of partners) Denominator: Total population of 15-49 year olds, disaggregate by sex and age groups 15-19, 20-24, 15-24, 25-49	UNGASS, PEPFAR GLIA WB scorecard
4. More than one sex partner in the past 12 months and reported using a condom during last sexual intercourse among men and women aged 15-49	**Women and men aged 15-49 who had more than one sex partner in the past 12 months and reported using a condom during last sexual intercourse**	Reported sex with more than one partner in the past 12 months and used a condom during last sex Denominator: Total population aged 15-49, disaggregated by sex and age groups 15-19, 20-24, 15-24, 25-49	UNGASS GLIA WB scorecard
5. Sex with a non-regular partner in the last 12 months among men and women aged 15-49	Percent of men and women aged 15-49 who reported having sex with a non-regular partner in the past 12 month	Reported sex with a non-regular partner in the past 12 months. Denominator: Total population aged 15-49, disaggregated by sex and age groups 15-24, 25-49	UNGASS GLIA WB scorecard

Indicator	Definition	Construction	International reporting
6. Condom use at last sex with a non-regular partner among men and women aged 15-49	Percent of men and women who say they used a condom the last time they had sex with a non-regular partner, of those who had sex with a non-regular partner in the last 12 months	Condom use at last sex with a non-regular partner Denominator: Men and women, aged 15-49 who had a non-regular sex partner in the past 12 months, disaggregated by sex and age groups 15-24, 25-49	GLIA WB Scoreccard
7. Sex with a transactional partner in the last 12 months among men and women aged 15-49	Percent of men and women aged 15-49 who reported having sex with a transactional partner in the past 12 month	Reported sex with a transactional partner in the past 12 months. Denominator: Total population aged 15-49, disaggregated by sex and age groups 15-24, 25-49	GLIA
8. Condom use at last sex with a transactional partner among men and women aged 15-49	Percent of men and women who say they used a condom the last time they had sex with a transactional partner, of those who had sex with a transactional partner in the last 12 months	Condom use at last sex with a transactional partner Denominator: Men and women, aged 15-49 who had a transactional sex partner in the past 12 months, disaggregated by sex and age groups 15-24, 25-49	UNGASS, PEPFAR GLIA
HIV testing			
9. Percent of men and women aged 15-49 received an HIV test in the past 12 months and know their results	Percent of men and women aged 15–49 who have been tested for HIV in the last 12 months and received their test results the last time they were tested	Tested for HIV in the past 12 months and received their results Denominator: Total population aged 15-49, disaggregated by gender, age groups 15-19, 20-24, 15-24 and 25-49	UNGASS, PEPFAR GLIA WB scorecard

Indicator	Definition	Construction	International reporting
STI health facility utilization			
10. Percent of men and women aged 15-49 who had an STI symptom in the past 12 months and sought treatment at a health facility	Percent of men and women aged 15–49 who report an STI symptom (genital ulcer or sore, unusual genital discharge) in the last 12 months and went to a public or private health facility as their FIRST recourse for treatment	First place went for treatment of STI symptom was public or private health facility Denominator: Total population aged 15-49 with an STI symptom in past 12 months, disaggregated by sex and age groups 15-24, 25-49	None
Knowledge, attitudes and misconceptions			
11. Percent of men and women aged 15-49 with comprehensive correct knowledge of HIV/AIDS	Percent of men and women who correctly identify two major ways of preventing HIV sexual transmission: Using condoms Limiting sex to one faithful, uninfected partner AND who reject the two most common misconceptions: Mosquitoes transmit HIV Sharing food with an infected person transmits HIV AND who know that: A healthy-looking person can transmit HIV	Composite indicator constructed from the 5 prompted knowledge and misconceptions questions. Person must respond correctly to all 5 questions. Denominator: Total population aged 15-49, disaggregated by sex and age groups 15-24, 25-49	MDG, Indicator 19b UNGASS PEPFAR GLIA

Indicator	Definition	Construction	International reporting
12. Percent of men and women aged 15-49 with accepting attitudes towards PLHIV	Percent of men and women who report that they would be willing to care for a family member sick with AIDS in their own household AND Would buy fresh vegetables from a shopkeeper with HIV AND think a teacher with HIV should be allowed to continue working AND Does not think that it should be kept a secret if a family member had HIV	Composite indicator constructed from the 4 prompted attitudes questions. Person must respond correctly to all 4 questions. Denominator: Total population aged 15-49, disaggregated by sex and age groups 15-24, 25-29	PEPFAR GLIA
Access to Prevention Programmes			
13. Percentage of men and women aged 15-49 who have been reached by HIV prevention programmes	Percent of men and women aged 15-49 who knew where they could receive and HIV test and had been given condoms in the past 12 months	Men and women reached by HIV prevention programmes Denominator: Total population of men and women 15-49, disaggregated by sex and age groups 15-24, 25-49	UNGASS, PEPFAR
Additional indicators for displacement situations			
14. Percent of women aged 15-49 who were forced to have sex in the past 12 months	Percent of women aged 15-49 who reported that they were forced to have sex in the past 12 months	Women forced to have sex in past 12 months Denominator: Total population of women	None
15. Percent of men and women residing in current community for 12 months or less	Percent of men and women aged 15-49 who reporting that they had resided in current community for 12 months or less	Resided in community for <=12 months Denominator: Total population aged 15-49, disaggregated by sex and age groups 15-24, 25-49	None

Indicator	Definition	Construction	International reporting
16. Percent of men and women away from home for four or more weeks in the past 12 months	Percent of men and women aged 15-49 who report that they had been away from home for four or more weeks in the past 12 months	Away from home for four or more weeks in the past 12 months Denominator: Total population aged 15-49, disaggregated by sex and age groups 15-24, 25-49	None
17. Percent of men and women who visit the surrounding community one or more times a month	Percent of men and women aged 15-49 who report visiting the surrounding community one or more times a month	Visit surrounding community one or more times a month Denominator: Total population aged 15-49, disaggregated by sex and age groups 15-24, 25-49	None

MDG	Millennium Development Goal
UNGASS	United National General Assembly Special Session on HIV/AIDS
PEPFAR	President's Emergency Plan for AIDS Relief
GLIA	Core indicator for the Great Lakes Initiative on AIDS
WB Scorecard	The World Bank has a standard set of indicators for all its projects, called an HIV scorecard
Non-regular sex partner	Person who is not a spouse or cohabitating sex partner or a transactional sex partner
Transactional sex partner	Person who was given or received money, gifts or favors in exchange for sex

Chapter 3

Survey flow

This chapter will provide you with an illustration of each of the steps involved in survey implementation, from the preparation stage up to analysis and dissemination.

Figure 1 below broadly outlines the activities involved in the following:

1. Survey preparation
2. Sampling
3. Questionnaire and informed consent
4. Survey documentation
5. Team composition and training
6. Data collection
7. Data management
8. Data analysis
9. Report and dissemination

This schematic is intended to give you an easy reference for the important steps in survey implementation. However, investigators will need to tailor their survey to their own timeframe and resources, and the steps may not occur in the exact order presented below.

For each of the steps in the survey 'flow' there is a chapter that discusses in-depth how they should be done.

Figure 1: Survey flowchart

Survey preparation

Collect *information* including previous surveys, census data, maps, population and household lists, local interventions

↓

Determine the *objectives* of the survey and core *indicators*

↓

Define the survey *population* and *geographic area*

↓

Determine *sampling* methodology and sample sizes

↓

Build partnerships with community leaders, UN and government agencies, NGOs, local authorities and generate *consensus* on survey objectives

↓

Communicate with refugee and host populations about the purpose of the survey

↓

Resolve *timing, access, participation and security issues*

↓

Obtain *ethical clearance*

↓

Define survey *team composition* and roles

↓

Define preliminary survey *timeframe*

↓

Develop a preliminary *budget*

↓

Sampling design

Finalize *sample sizes*

↓

Validate *accuracy of population data* and fill any information gaps

↓

Finalize *sampling design*

↓

Determine *number of field staff and days* needed to achieve sample size

↓

Refine the *budget and timeframe*

↓

Questionnaire design and testing

Select *core or expanded questionnaire*

↓

Carefully determine and make any necessary *adaptations* to questionnaire and consent form

↓

Translate into all locally relevant languages

↓

Field test the documents

↓

Revise the questionnaire and consent form based on field test results

↓

Back-translate all documents

↓

Finalize and *print* the questionnaire

↓

Survey documentation

Develop or adapt the *survey protocol* for local field conditions

↓

Develop or adapt *additional fieldwork forms*, such as household information sheets

↓

Determine the step-by-step *process of data collection* including *roles and responsibilities* of the team

↓

Document each step in the process of data collection, management and analysis in an *operational manual*

↓

Finalize the *timeframe* and *budget*

↓

Team composition and training

Determine the *field team composition* and *qualifications*

↓

Recruit necessary field staff and supervisors

↓

Train all field staff and supervisors, including mock data collection activities

↓

Review abilities of each member of the team and *select final members*

↓

Data collection

Finalize a detailed *schedule* for data collection

↓

Recruit participants

↓

Set up sites and conduct *interviews*

↓

Conduct *quality control* checks at each stage

↓

Collect fieldwork forms each day and store in a locked place

↓

Team *debriefing*

↓

Data management

Develop *data entry program* with quality control checks

↓

Train all data entry clerks on the questionnaire and data entry procedures

↓

Develop data entry *schedule*

↓

Enter all questionnaires and other field work *information* into the data set

↓

Clean all data sets rigorously prior to beginning analysis

↓

Combine and/or rename data sets if needed

↓

Collect and store questionnaires and fieldwork forms in a central, secured location upon completion

↓

Data analysis

Update *coding and analysis guide*

↓

Conduct analysis of *proportions and means*

↓

Conduct analysis of *core indicators*

↓

Review analysis for accuracy

↓

Document all results in results templates

↓

Conduct *additional analysis* of important findings and document results

↓

Report and dissemination

Share findings with key collaborators and population members to develop *conclusions and recommendations*

↓

Determine best *methods for disseminating findings* to important audiences and other ways of *using the data*

↓

Document entire survey process and the results in technical report and other mediums

↓

Disseminate findings widely to local and international partners, population members and decision-makers

Chapter 4

Survey preparation

While there are many technical aspects in organizing a BSS, the administrative and community preparation needs of the survey should not be overlooked, and the time and resources allocated to them should not be underestimated. Strong pre-survey groundwork with the necessary administrative support will be one of the determining factors for whether you are able to collect high quality data in a manner that is beneficial to both the participating populations and the agencies that will be the ultimate end-users of the information. This chapter will discuss each of the steps, both technical and administrative, that must be finalized in order to begin data collection. Some of the steps, such as deciding on survey objectives and designing a sampling frame, are covered in more depth other chapters.

Step 1: Collect information including previous surveys, census data, maps, population and household lists, national surveillance and other survey methodologies, and local intervention strategies

You should be collecting and reviewing all relevant information and literature while you are determining your survey objectives, populations, and coverage areas. Frequently other surveys have already been conducted that provide information on the populations of interest, methods, and lessons learned. The data in these reports can also assist in fine-tuning sample size calculations and methods (see Chapter 5: Sampling). Contact government offices, NGOs, and research organizations for past surveys. National undertakings such as the Demographic and Health Survey (DHS), which sometimes includes an HIV component, and the Reproductive Health Survey will often have disaggregated data for your region of interest.

The Ministry of Health may also have a national HIV surveillance system and documented survey methods. When it is appropriate you should adapt your survey approach so it is complementary to that of the nation-wide surveillance system. This will assist you in comparing the data to other areas of the country as well as providing valuable information to the host government.

Collecting all available population data, particularly census data and lists of administrative units in the survey catchment site, will be crucial to developing your sampling methodology and defining your population. And understanding what HIV programs are planned or already in place in your area will assist in finalizing your survey objectives and developing locally relevant questions on exposure and access to interventions (see Chapter 6: Questionnaire and Informed Consent).

Step 2: Determine the objectives and core indicators

As discussed in *Chapter 2: BSS objectives and core indicators*, surveys that measure behavioral risk in a population have a number of objectives that are common across most studies. However, before designing a BSS for your situation, these objectives must be reviewed, and it should be determined whether there are additional objectives and indicators that are unique to your own survey site and/or population. Chapter 2 discusses how to go about this process in more detail.

Step 3: Define the survey populations and geographic coverage areas

While these surveys are designed to be conducted among refugees and other displaced groups and their surrounding populations, exactly how these populations are defined and within what geographic boundaries they are to be recruited will be different for every situation. Some questions that will need to be considered in making these decisions include - but are not limited to - the following:

What population constitutes the 'refugee community'? Do they live in one camp? Or is the area of interest made up of several camps? Are there defined boundaries for the area? Do you want to include all or only a portion of the refugee community? If refugees have been living in the area for a long period of time, some perhaps even having been born there, they may not consider themselves refugees. Is this sub-group going to be included in the survey? Does the government recognize all nationalities in the camp as having official refugee status? If not, will you include those nationalities in the camp that are not recognized as refugees? Do the HIV interventions in the area cover all nationalities or other sub-groups of refugees? If not, will you include the groups that will not receive interventions? Are there any factors that may affect the ability to recruit some refugee participants? What constitutes a 'member of the household'? What will be the selection criteria for the refugee population?

What population constitutes the 'surrounding community'? Are there villages that are clearly contiguous to the camps or is the national population dispersed across a wide area? How far away from the camp constitutes a 'surrounding population'? Do the HIV interventions in the area cover all nationals? If not, will you include the groups that will not receive interventions? How does the Ministry of Health define their population for national HIV surveillance, and how will that affect your population choices? Are there any factors that may affect the ability to recruit some national participants? What constitutes a 'member of the household'? What will be the selection criteria for the surrounding community?

After posing each of these questions, the population definition and coverage area of the survey must be clearly documented, as shown in the example below.

Example from Uganda of selection criteria for populations and geographic coverage areas

In Uganda, two areas were selected for the BSS in refugee and national communities. The initial areas, Mbarara and Hoima, were chosen because they would be the sites for HIV interventions in both the refugee and surrounding communities, and data were needed to design

appropriate interventions and evaluate program progress. In Mbarara, there are two refugee settlements, Nakivale and Oruchinga. These settlements are adjacent to each other and interventions were planned for the two areas, so both were included in the survey. In Hoima, there is only one refugee settlement, Kyangwali, so only this site was included. The UNHCR offices in both areas had already defined borders for the settlements, so all refugees falling within this area were included.

Because the settlements in both areas were dispersed and there were large distances between the villages where refugees resided, determining the area that made up the 'surrounding community' was more difficult than identifying the refugee area. One of the objectives of the survey was to determine frequency of interaction between the refugee and host populations, so the national population was defined as only those residing in villages within two kilometers of the refugee population.

Even though people displaced from Rwanda no longer had official refugee status in Uganda, all sub-populations of refugees and nationals within these geographic perimeters would be involved in the HIV interventions (regardless of status), so all nationalities were included in the survey. Many of the refugees living in the sites had been there for more than a decade and some had Ugandan citizenship through birth or previous marriage. However, the refugee population was defined as those living within the pre-determined borders of the refugee settlement in a home where the head of household was a refugee, and not those who self-defined as either 'refugee' or non-refugee'. In addition, some people living in the villages surrounding the settlements were former refugees who had married host county nationals and become Ugandan citizens. Again, the definition of a participant from the 'surrounding population' was determined by where the person resided and the refugee status of the head of household.

The primary objective of the survey was to measure behavioral risk among refugees and surrounding populations between the ages of 15 and 49, disaggregated by specific age bands. However, the Ugandan national HIV and behavioral surveillance system measures similar indicators for the age range of 15-59 years old. In order to conform to the population definition of the host Ministry of Health, the survey population was extended to include all men and women between the ages of 15 and 59.

After defining the survey populations and geographic boundaries as well as reviewing the national HIV surveillance guidelines, the population inclusion criteria in the two Ugandan survey sites were the following:

- All men and women between the ages of 15 and 59 were included
- Members of the household were defined as people who had been living in the household for at least two weeks and shared meals in the home
- Refugee status was determined by the location of the household and the status of the head of the household
- Refugees were those residing within the UNHCR-defined boundaries of the settlement
- Nationals were those residing in a village within a two kilometer proximity of a refugee village

Stop. Have you reviewed all relevant surveys and population data that are already available? Have you determined your survey objectives and what you want to measure? Do you have a clear definition and inclusion criteria for each of the survey populations? Have you defined the geographic coverage area of the survey? If you have not done this, do not move on to the next step in survey preparation.

Step 4: Determine survey sampling methodology and sample size

The mechanics of calculating sample sizes and creating sampling frames for BSS are explained in detail in Chapter 5. However, sampling strategies must be addressed during the survey preparation phase and should never be left until just before the survey goes into the field. The sample size of your survey will be determined in part by what type of sampling method you employ, and it could be different for each population and area. Sample methods may range from a technically simpler (and less costly) simple random sampling from complete household lists to a more complex cluster sampling methodology.

All sampling decisions will need to be reflected in the timeframe and – very importantly – the budget of your survey. It is not feasible, or statistically rigorous, to let your budget determine you sample sizes or sampling methods. Instead, you need to decide what sample size and sampling methods will be necessary to achieve the objectives of your survey, and then you must be prepared to provide enough funds to cover these needs. If your budget cannot cover the staff, time and logistical costs necessary for proper sampling, then you should reconsider whether conducting a survey is advisable or redefine your survey objectives. Cutting costs will compromise your sampling and, in turn, the accuracy and usefulness of your results. In order to account for all costs, sampling must be considered very early in the survey preparation process.

Step 5: Build partnerships with community leaders, government agencies, NGOs, and local authorities to generate consensus on survey objectives and implementation plans

Behavioral surveillance surveys are about more than just an agency with an idea and a lead researcher hired to manage the project. The surveys can provide much needed information for the local community on HIV risks and vulnerability. But if not carefully planned they can also be disruptive and resource intensive for both the people who are asked to participate and the organizations that assist in their implementation or otherwise come into contact with the survey teams.

Organizations including local NGOs, agencies and government offices that will be participating in the surveys by providing logistical assistance, field support, or other services must be brought into survey planning as early as possible. Not only should they understand and agree with the objectives of the survey, but they must also fully comprehend the amount of time and staff resources needed to successfully complete the exercise and agree to take re-

sponsibility for their part in making it happen. In order to achieve this, it is very useful to create a local survey advisory board of key people who meet regularly. If all partners understand and buy into the survey objectives and process during the preparation phase, the survey will be easier to complete and the end users of the data will be better able to understand and apply the results to their work.

Throughout survey planning, open communication with survey partners and other local key stakeholders will also help identify potential difficulties at a stage where they can be more easily managed. Questions that should be discussed in partnership with local stakeholders are presented in Step 7 of this chapter

Step 6: Communicate with displaced and host populations about the purpose of the survey

People will be understandably wary of an unannounced group of strangers coming to their door to ask them sensitive questions. Some populations may have already have participated in other surveys. They may already be familiar with the process of providing informed consent and answering a questionnaire which can assist the interviewers. Or they may be tired of spending their time answering questions without any discernable benefits to themselves. Some populations may never have taken part in a survey and may think their participation will influence access to food or other services, regardless of how clearly the informed consent states otherwise.

The objectives of the survey need to be communicated clearly to the communities well before the survey team arrives in their villages. This process begins with sensitizing village leaders, but it does not end there. Not all leaders may be available for or willing to take part in a community meeting. Organizers, preferably those who are already known and trusted in the area, should make an effort to go directly to the leaders as well as the community members to discuss the survey and to clarify its objectives. Often information given to a village leader will not trickle down to the residents of the community. Use opportunities when many members of the community will gather in one place, such as during village meetings or health outreach programs, to discuss the survey. But be very careful to ensure that the communities do not perceive participation in the BSS as linked to their access to other services.

Step 7: Address survey timing, population access, community participation, and security issues

Below is a list of questions to be addressed while planning and budgeting for your field work. This is only a starter list and is not exhaustive. You should also consider other issues that may be unique to your local situation.

Data collection timing

- When is the rainy season? When are local festivals? What other events, such as market days, should be considered in organizing the data collection schedule?
- Is it the planting or harvesting season? Are people likely to be working in their fields during the day? What is the best time of the day to recruit people? Remember, revisits

to households in search of absent members are costly and time consuming. It is easiest if data collection is scheduled so that a maximum number of household members will be present during the first visit.

- Is there a time of day or a day of the week when it is more common for people to be consuming alcohol? While this may not be important in many survey locations, in localities where alcohol consumption is more prevalent, you will want to avoid recruiting during times when people may be intoxicated.
- During what days and in what areas do food distribution, registration, or other activities take place? How can data collection be organized to ensure the least amount of disruption to vital programs?

Access to villages

- What is the condition of the roads in the survey catchment area? Are there any roads that are impassable? Are 4 wheel drive cars necessary? Are 4 wheel drives available for hire or on loan?
- What is the distance between villages that will be sampled? What is the distance between the villages and your survey team's base of operations?

Community participation and expectations

- Do key stakeholders and community members understand the objectives of the survey? Are they willing to participate? Are there any sub-populations in the community which may require additional sensitization?
- What was the community's response to other surveys that have taken place in the area?
- Have participants been promised things (such as interventions) by other researchers that have not been seen through?
- Was confidentiality respected by the other research teams?
- What types of incentives (particularly cash) were provided by other researchers?
- How may these issues affect participation? How can problems encountered by previous research activities be minimized and how can this survey activity learn from past experiences?

Security

- What is the security situation in the camps and surrounding villages? Is it likely to change?
- Are there any areas included in the survey that are less secure and where local team members and/or international staff persons should not travel?
- Is it safe to travel at night? How late can survey teams work so they can still get home before dark? How many hours will you have left each day for data collection? Do not put the personal safety of your teams at risk in order to complete the survey faster. Be conservative and extend the time allocated for data collection if necessary.
- What security precautions are necessary? Is there mobile phone access in the area? Are radios available? Are the vehicles in good shape? Do they have spare tires?

Step 8: Define survey team composition and members' roles and responsibilities

In Chapter 8, the composition and training of the field team is discussed at length, but the chapter focus is on the data collection team. While the interviewers and supervisors are important members of your survey team, there are other people who need to be included to ensure the successful completion of the project.

1. **Lead investigator** – One person should have overall responsibility for the coordination and implementation of the BSS. This person should be an experienced researcher, preferably one who has conducted previous quantitative behavioral surveys. The lead investigator should be empowered to make all final technical and administrative decisions, though this should be done in consultation with other core staff.
2. **Administrative manager** – One person should be responsible for the completion of all administrative aspects of the survey. This includes managing survey schedules, budgets, contracts, payments, car rentals/loans, drivers, petrol, computers, printing and photocopying, and space for training, data entry and other activities. This person may either handle the work themselves or supervise other staff completing it, such as a **logistics manager** and **finance staff.** The lead investigator cannot also manage the administration of the survey because it is very time intensive and requires a high level of local knowledge. Ideally, the administrative manager would be a staff member of the implementing agency. However, BSS preparation and implementation is a full-time job. The administrative manager must have dedicated time and should not be asked to balance survey responsibilities on top of their other workload.
3. **Data manager** – Frequently, data entry is done at the same time as data collection, with questionnaires being inputted immediately after they return from the field. One person should have responsibility for collecting the questionnaires from the supervisors, organizing and tracking all returned questionnaires and other fieldwork forms, managing data entry clerks, checking data entry quality daily, and collating and managing the raw data sets. Again, this is a full time job, and the lead investigator **cannot** manage both the field work and the data entry at the same time. The cost of hiring a data manager may ultimately be far less than the time and budget required for cleaning, and possibly re-entering, datasets that are poor quality.
4. **Data entry clerks** – The number of data entry clerks you hire will be dependant on the sample size, number of days allocated for data entry and management, number of computers that are available, number of qualified people that can be recruited, available space, and whether you do double data entry. An experienced data entry clerk should be able to input 50 BSS questionnaires a day depending on the length of the survey. However, it will take some time for the clerks, regardless of experience, to become comfortable with the questionnaire and data set format. You should expect that they will take a number of days to get up to full speed, and many will never reach that number.
5. **Data collection supervisors and interviewers** – The number of data collection staff you need to employ is discussed at length in *Chapter 8: Team composition and training*.
6. **Field guides** – If the members of your team are not very familiar with the administrative areas where the survey is being conducted, field guides who are part of the local communities should be recruited to assist them during data collection.

7. **Drivers** – A vital survey role is that of the driver. The number of drivers you will need will depend on how many cars are available, what distance they need to travel each day, how many field teams can be transported in each car, and how many days of fieldwork are scheduled. If the drivers you use are staff members or on-loan from other agencies, it should be recognized that they will need to stay with the field teams for the full day. If cars and drivers are hired, remember that managing the drivers – along with contracts, payments, car maintenance, and petrol procurement – is a very time consuming job.
8. **Data analyst** – The person responsible for the final cleaning and analysis of the data sets, as well as the report, is frequently the lead investigator. If this person does not have the necessary skills, then a data analyst needs to be hired. Preferably, the data analyst will also play a role in data collection so that he or she has a strong understanding of the methods and implementation process of the survey.
9. **Translators** – You will often be working in a number of different languages, and not all field staff and trainers will be able to communicate in all languages. Professional translators should be hired (and trained) for translating the questionnaires. You may also need them for the interviewer and supervisor training.

Step 9: Develop the preliminary survey timeframe

In order to clearly define the full scope of your survey and effectively budget for all activities, you need to lay out every step necessary to conduct the BSS and the timeframe in which it will be done. Each activity should also indicate the number of days and number of staff necessary for its completion. While schedules frequently change, it is still useful to develop a draft timeframe early in the preparation stage.

An example timeframe is provided below. An optimal number of days is given for each survey step, though these may vary depending on how many languages you are working in, the size of your field teams, sample size, number of populations, and many other considerations.

Stop: Have you determined your sample size and sampling methodology, built partnerships and sensitized the community, resolved issues related to survey implementation, determined team composition and number of staff needed, and developed a timeframe? If you have not done this, do not move on to the next step in survey preparation.

Step 10: Develop a realistic budget

While you may already have allocated a certain amount of money to conduct the BSS, you will not know the true costs of the survey until you have completed all of the previous steps. Even then, you should always be prepared for field contingencies such as adverse weather conditions that delay data collection, having to hire additional cars, changes in petrol prices, computer crashes, and the like.

The budget should reflect salaries and/or per diems determined by the number of days and staff outlined in your survey timeframe. Additional costs that should be built into your budget include, but are not limited to:

- Renting space (training, data entry, other needs)
- Printing and photocopying
- Office supplies such as pens, notepads, Post-It notes, colored tags (handy for marking questionnaires), easel paper, folders, staplers, paperclips
- Communication including SIM cards for mobile phones, radios, e-mail, local and international calls
- Computer rental
- Software
- Car rental for data collection and for travel between survey sites
- Car maintenance
- Fuel costs
- Accommodation, when necessary
- Participant incentives, if appropriate
- Supplies for field staff such as folders to carry questionnaires and other documents, umbrellas to keep off rain/sun, rain boots and raincoats, strong shoes if long distances have to be covered
- Name tags to identify survey members and their affiliation to local leaders and participants
- Meals during training and other activities
- Certificates

Prior to beginning your BSS, the survey protocol and questionnaires should be approved by the appropriate ethical review board(s). Usually the country in which you are working will have a review board, often housed in the Ministry of Health or other government body, and often the funding and/or implementing agency will also have ethical review requirements. Identify the process and timeframe for obtaining ethical clearance as early as possible in the survey design stage. It may take weeks, and not infrequently months, to submit the necessary documents, respond to inquiries from the board, and obtain full clearance. It is mandatory that, when an ethical review is required, it is completed satisfactorily before data collection begins. *Chapter 13: Research ethics* also discusses some of the key issues in conducting an ethical survey.

Table 2: Example of preliminary BSS timeframe

Activity	Week 1	Week 2	Week 3	Week 4			
Survey preparation							
Review relevant material							
Finalize objectives and indicators							
Define populations and geographic coverage areas							
Determine sample size and methodology							
Build partnerships with local leaders and organizations							
Sensitize community							
Finalize staff numbers and roles and responsibilities							
Develop preliminary timeframe							
Develop budget							
Obtain ethical clearance							
Questionnaire	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11
Adapt questionnaire and consent form							
Translate questionnaires		5 languages = 5 translators * 3 days each					
Field test questionnaires							
Back translate questionnaires					5 languages = 5 translators * 3 days each		
Finalize and print questionnaire							
Sampling frame							
Develop sampling frame and select households/clusters							
Survey documentation							
Develop survey protocol							
Develop operational manual							
Finalize all field and training documents							
Training and logistics	Week 12	Week 13	Week 14	Week 15	Week 16	Week 17	Week 18
Final recruitment of supervisors and interviewers	On-going prior to training						
Organize training							
Train field staff		5 days * 33 staff and 3 translators					
Finalize logistics for data collection							
Data collection							
Field work			33 staff and 5 drivers * 15 days				
Data management							
Double data entry			6 clerks and one supervisor * 30 days				
Data cleaning							
Data analysis	Week 19	Week 20	Week 21	Week 22	Week 23	Week 24	Week 25
Final cleaning and coding							
Data analysis							
Report							
Data dissemination							
Dissemination	On-going after results are completed						

Survey documentation including developing the protocol and operations manual is another vital step in the survey preparation process. Because of its importance it is discussed separately in *Chapter 7: Survey documentation*, but it should not be overlooked while preparing for your survey.

Checklist 1: Survey preparation

- ☐ Collect and review all relevant reports and population data
- ☐ Determine the objectives and core indicators
- ☐ Define the survey populations and geographic coverage areas
- ☐ Determine the sampling methodology and sample size
- ☐ Build partnerships with local leaders and organizations
- ☐ Communicate survey objectives to communities
- ☐ Resolve issues regarding survey timing, community access and participation, and security
- ☐ Determine staff needed and define roles and responsibilities
- ☐ Develop preliminary timeframe
- ☐ Develop a realistic budget
- ☐ Obtain ethical clearance from the appropriate review board(s)

Chapter 5

Sampling

There are a number of excellent resources that have been developed that provide guidance for sampling in population-based surveys. The following chapter excerpts some of the resources and provides additional information that is specific to surveys among displaced populations and their surrounding communities.

Definition of sampling terms

The following is excerpted from *Behavioral Surveillance Surveys: Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV*, Family Health International, 2000, http://www.fhi.org/en/HIVAIDS/pub/guide/bssguidelines.htm

Excerpt

- **Sampling universe:** The entire group of sampling units (commonly households or persons) who are eligible to be included in the survey sample. This population should match the population for which you are trying to estimate the survey outcomes. For example, if a survey is designed to measure HIV behavioral risks among a refugee population and compare them to risks in the surrounding community you have two sampling universes: (1) every eligible person in the entire refugee community (2) every eligible person in the entire host community residing in the geographic survey area.
- **Sampling frame:** The list of all the sampling units from which you will choose your sample. The sampling frame should match the sampling universe as closely as possible, however sometimes it will be slightly different. For instance, if there are insecure areas within the sampling universe where it would be dangerous to take a study team, the people residing in these areas may not be included even if they are part of the survey population.
- **Sampling unit:** The unit that is selected during the process of sampling. If you select households from a list of all households in the population, the sampling unit is the household. If you are selecting villages in the first stage of cluster sampling, the sampling unit at the first sample stage (or primary sampling unit) is the village.
- **Basic sampling unit or elementary unit:** The sampling unit at the last stage of sampling. In a multi-stage cluster survey, if you first select villages, then select households within the village, the basic sampling unit would be the household.
- **Sampling fraction:** The proportion of all sampling units in the sampling frame which were selected or will be selected for the sample. If you want a sample of 100 households from a list of 10,000 households, the sampling fraction is 100/10,000 = 0.01. Thus, one of every 100 households in the sampling frame is selected for the survey sample.

- **Sampling interval:** The inverse of the sampling fraction. This is used when carrying out systematic random sampling. In the example above, the sampling interval would be 100. If you are selecting households by walking down a street and selecting every 100th household, you would count households until 100 is reached and include that household in the sample. Then you would continue counting until the 200th household is reached and include that household in the sample, etc...
- **Respondent:** The person who answers questions during an interview.
- **Design effect:** Measurement of how evenly, or unevenly, the outcome is distributed in the population being sampled. For example, if you think knowledge of HIV prevention is about the same in all parts of the population, then the design effect is probably low. If you think that the prevalence of transactional sex is higher in some areas of the population than in others (ie people living near a main road), then the design effect for this outcome would be higher. Frequently the design effect is not known, and can range significantly between different outcome indicators.

End of excerpt

Calculating sample size

The sample size required for your survey will be based on a number of different factors, each of which will need to be considered for your individual objectives and circumstances. You cannot chose a number to be your sample size based simply on what other surveys in your area have used or the size of your budget. The sample size will affect the precision with which you are measuring the key indicators in the survey and your ability to measure trends in behavior over time, so it must be tailored to your individual survey and specific context.

In order to determine your sample size, you must go through each of the following steps.

Step 1: Define your sampling universe.

What populations are you including in the survey? Do you want to be able to compare each of the populations separately, or are you going to group them together during the analysis? What geographic areas are you including in the survey? Do you want to be able to compare each of the geographic areas separately, or are you going to group them together during the analysis? If you are surveying refugees in several camps that are close to each other, do you want to look at behavioral risk factors separately for each of the camps, or are you going to combine all of the camps together for your survey? If you have two areas of the country with large numbers of displaced people, do you want to analyze their results separately or will you aggregate all the data together as a single population group?

The definition of your sampling universe will depend on the objectives of your survey. If a key objective of your survey is to describe risk factors of different refugee populations residing in three different areas of the country, then you will have three separate sampling universes for the refugees. If another objective of your survey is to compare the behavioral risks and access to interventions among the three refugee populations and their surrounding commu-

nities, the host populations will be three additional sampling universes. In total you would have six separate samples in your survey, meaning the sample size will be applied to each of these populations separately. You cannot calculate a sample size and spread it across all of the populations in your sample or you will lose the power to analyze the different refugee and host population groups separately and compare the results.

Example from Uganda: In Uganda, there were two areas of the country where HIV programs were being designed for populations in refugee settlements and their surrounding communities. The survey objective was to measure the differences in behaviors and access to interventions between the two different refugee areas and to also measure the difference between the refugee populations and the surrounding communities. In one survey area, there were two refugee settlements abutting each other. Both settlements were considered to be similar in composition and characteristics and both were to receive the same HIV interventions. These settlements were combined together as a single survey population with the understanding that the results from the two settlements would not be analyzed separately. In total, there were four distinct populations, or sampling universes, that were included in the survey.

Step 2: Determine the key indicators that you want to measure, with what precision you want to measure them, and if you will be evaluating changes in these behaviors over multiple survey rounds (i.e. trends).

While budgets should not be the determining factor in calculating the sample size of your survey, restrictive budgets are a reality for any organization working in HIV prevention. For this reason, you must understand what indicators your sample size will allow you to measure more precisely, and where you will have to sacrifice tight confidence intervals for a wider and less precise estimate of your indicator.

Measurements of behaviors at a single point in time require a smaller sample size than measuring changes in these behaviors between multiple survey rounds. If an objective of your BSS is to measure behavioral trends over time, then your sample size will be somewhat larger than if the survey is intended to be a one-off description of your population.

If you are measuring behavioral trends, you will then have to determine what the baseline estimated prevalence of the specific behavior you are measuring is at the time of your survey and what magnitude of behavioral change you wish to measure and compare. Previous research in the region among similar populations can provide you with reasonable estimates of the baseline behavioral prevalence. The minimum percentage increase – or decrease – in a behavior that you want to measure should be determined based on previous research and program targets. When this information is unavailable, it is common to use a 10% or 15% change between survey rounds for behavioral indicators, but this can vary depending on the indicators of interest, the amount of time between surveys, and the intensity of interventions in the area that will be affected the change. What this means is that your sample size will be sufficiently large to measure, for instance, a 10% or larger change in the behavior from baseline prevalence, but it will not be able to measure statistically significant differences between survey rounds that are smaller than 10%. The smaller the percentage of change

that you wish to measure between survey rounds, the larger your sample size will need to be (see the table below for illustrative sample sizes).

Example from Uganda: The BSS in Uganda was designed to measure the same objectives and key indictors described in the earlier chapters of this manual, but not all indicators were feasible to use in calculating the sample size. For instance, transactional sex and casual sex partners among the male and female youth populations are two important behavior risks to estimate, however, they were already known to be infrequent activities among the youth populations. While the sample size of the survey was adequate to estimate the prevalence of these behaviors among youth at baseline, it was not possible to achieve a sample size that would be large enough to identify statistically significant changes of 15% or less in these behaviors over time.

Instead, the indicators selected for calculating the sample size in the Uganda BSS in order to measure trends over time were the following:

1. More than one sex partner in the past 12 months among men and women aged 15-49 years
2. Comprehensive correct knowledge of HIV/AIDS among men and women aged 15-49 years
3. Accepting attitudes towards PLHIV among men and women aged 15-49 years

The table below provides an illustration of the sample sizes necessary to measure changes in the prevalence of behaviors over time.

Sample sizes necessary to calculate change from baseline prevalence over time*

Baseline prevalence	Sample size*			
	5% change	10% change	15% change	20% change
10%	540	156	78	48
20%	860	231	108	64
30%	1083	280	128	73
40%	1206	305	136	76
50%	1231	305	133	73
60%	1157	280	119	64
70%	984	231	95	48
80%	712	156	59	27

* Sample size assumes 95% alpha, 80% beta, no design effect and no non-response

Step 3: Determine what type of sampling methodology you will use and whether you will have to increase your sample size to account for design effect.

Different sampling methodologies for BSS surveys will be discussed later in this chapter. Before finalizing the sample size for your survey, you will have to determine what type of sampling methodology you will use. If you employ a multi-stage cluster sampling methodology, you will have to inflate your sample size by a certain factor to account for the effect the survey design will have on the measurement of standard error, called 'design effect'. Frequently, the design effect is not known at the outset of the survey and a standard factor of 2, or sometimes less, is applied. The more homogenous your population, the lower your potential design effect will be. Before determining the design effect for your sample size, you should review literature and reports of the design effects used in other cluster surveys done in your populations and regions relating to the specific indicators you wish to measure. More detailed information on design effect for BSS in high risk and vulnerable popuations is provided in Chapter 11: Data analysis and in *Behavioral Surveillance Surveys: Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV*, Family Health International, 2000.

Example from Uganda: No complete household lists were available in Uganda for either the refugee or surrounding populations and the households were spread far apart. Therefore, households were selected for the survey using a cluster sampling methodology. Because a number of different ethnic groups were in the survey and there was a high potential of heterogeneity among respondents from different countries, a design effect of 2 was determined to be adequate to account for cluster sampling, and, thus, the sample was doubled in size.

Step 4: Calculate your sample size.

To determine your sample size, you will need the following information

1. Are the results going to be used to measure trends across multiple survey rounds or only a single round?
2. The indicators of interest you want to measure
3. The estimated baseline prevalence in the indicators of interest
4. The % change you want to be able to detect in behavior to be measured (for surveys measuring trends)
5. The estimated non-response rate (including refusals and absences)
6. The estimated design effect (for cluster surveys)
7. The desired level of statistical significance
8. The desired level of statistical power

Sample size calculation comparing two different survey rounds

The following formula can be used to calculate sample sizes for a BSS designed to measure trends in behaviors over time (excerpted from *Behavioral Surveillance Surveys: Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV*, Family Health International, 2000).

$$n = D\,\frac{\left[\sqrt{2P(1-P)}\,Z_{1-\alpha} + \sqrt{P_1(1-P_1) + P_2(1-P_2)}\,Z_{1-\beta}\right]^2}{\Delta^2}$$

Where:

D = design effect;
P_1 = the estimated proportion at the time of the first survey;
P_2 = the proportion at some future date such that the quantity (P_2 - P_1) is the size of the magnitude of change it is desired to be able to detect;
$P = (P_1 + P_2) / 2$;
$\Delta^2 = (P_2 - P_1)^2$;
$Z_{1-\alpha}$ = the z-score corresponding to the probability with which it is desired to be able to conclude that an observed change of size (P_2 - P_1) would not have occurred by chance;
$Z_{1-\beta}$ = the z-score corresponding to the degree of confidence with which it is desired to be certain of detecting a change of size (P_2 - P_1) if one actually occurred; and
$\alpha = 0.05$ ($Z_{1-\alpha} = 1.65$) $\beta = 0.20$ ($Z_{1-\beta} = 0.84$).

The following table provides estimated sample sizes for surveys that measure trends over time for both simple random sampling and cluster sampling. In this example, the non-response rate is estimated to be 10% and the design effect is 2.

Baseline prevalence	Sample size							
	5% change		10% change		15% change		20% change	
	SRS	Cluster	SRS	Cluster	SRS	Cluster	SRS	Cluster
10%	600	1198	174	348	87	174	54	107
20%	956	1912	256	512	241	120	71	142
30%	1203	2406	311	622	142	284	81	162
40%	1340	2680	338	677	151	302	85	170
50%	1368	2735	338	677	148	296	81	162
60%	1285	2571	311	622	133	266	71	142
70%	1093	2186	256	512	105	210	54	108
80%	791	1583	174	348	66	131	30	59

Sample size calculation for a single survey

There are a number of helpful on-line resources to assist you in calculating the sample size for your survey. For more information, please refer to the following:

Measuring and Interpreting Malnutrition and Mortality: A Manual, US Centers for Disease Control and Prevention and the World Food Programme, 2005
http://www.unhcr.org/publ/PUBL/45f6abc92.pdf

Food and Nutrition Technical Assistance: Sampling Guide, http://www.fantaproject.org/publications/sampling.shtml

Behavioral Surveillance Surveys: Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV, Family Health International, 2000
http://www.fhi.org/en/HIVAIDS/pub/guide/bssguidelines.htm

Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations: The SMART Protocol, January 2005, http://www.smartindicators.org/SMART_Protocol_01-27-05.pdf

Sample size calculator for cross-sectional surveys, freeware from Emory University,
http://www.sph.emory.edu/fficdckms/Sample_size_for_comparing_two_cross-sectional_surveys.html

SampleSX, sample size calculator for cross-sectional surveys, freeware from Brixton Health,
http://www.brixtonhealth.com/index.html
Note: *This tool does not calculate sample sizes for surveys measuring changes over time.*

Choosing your sampling methodology

The following is excerpted from *Behavioral Surveillance Surveys: Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV*, Family Health International, 2000
http://www.fhi.org/en/HIVAIDS/pub/guide/bssguidelines.htm

Excerpt

Probability and non-probability sampling methods: issues and challenges

Sampling procedures may be thought of as falling into two broad classes: formal or probability methods, and informal or non-probability methods. In a *probability sample*, every person in the defined universe may be selected into the sample, with a known (non-zero) probability. Probability sampling tends in practice to be characterized by the use of lists or sampling frames to select the sample, and by clearly defined sample selection procedures. With a probability sample, it is possible to use the data themselves to estimate the *sampling error*, or the effect of random fluctuations in sample selection on the accuracy of the observed results. Estimates of population characteristics derived from surveys based upon probability sampling

methods may be expected to approximate the "true" population value (i.e. proportion or mean) within a specified margin of error with a known probability.

Under the heading of *non-probability sampling methods* are a variety of approaches that are not based upon the statistical principles which govern probability samples. There are various reasons for using non-probability methods. Some methods (e.g. *snowball or network sampling*) are designed for use when probability sampling is not feasible. In snowball sampling, key informants in a sub-population identify other members of their community. These people are contacted, and they in turn identify further contacts. The process goes on until an adequate sample is achieved. Other methods (e.g. purposive sampling) are designed to provide the maximum amount of information possible for key groups of study subjects in order to develop and/or test social theories. Yet others (convenience sampling) are designed to obtain a sample of subjects at the least possible cost. In general, non-probability sampling methods are not intended to produce "representative" data for larger populations, although they are sometimes (incorrectly) used to try to do so.

Probability sampling has two major advantages. Firstly, it is less prone to bias than non-probability methods and secondly, it permits the application of statistical theory to estimate sampling error from the survey data themselves. Consistent use of probability sampling methods in the context of BSS has the critical advantage of producing data which are comparable from one survey to the next, and which can therefore be used to measure statistically significant changes in risk behavior over time. Therefore probability sampling methods are the preferred choice for BSS whenever feasible. The major disadvantage of probability sampling is that a list or sampling frame is needed, and this can take time and resources to produce. While there are ways to make the task of developing sampling frames less costly and time consuming, the use of probability sampling methods will nevertheless involve greater time and expense than sampling approaches that do not require a list or sampling frame.

While they are generally cheaper and easier to use, non-probability sampling methods have several important drawbacks. The first is the risk of *sampling bias* resulting from the subjectivity that often enters into the sample selection process. Where a list of sampling units is not available from which to select a sample following fixed rules, thee is the danger that certain types of subjects will be disproportionately included and others disproportionately excluded from the sample. Secondly, there is the issue of reliability, which is of key importance for surveys intended to monitor behavioral trends over time. Where sample selection criteria are not defined in operationally precise terms so that they can be replicated in subsequent survey rounds, there is a danger that observed changes will be due to changes in sampling rather than real changes in behavior. Finally, non-probability methods provide no statistical basis for assessing the precision or reliability of survey estimates. In fact, conventional statistical tests cannot be used reliably with non-probability samples, although in practice this limitation is often overlooked.

In the end, the issue boils down to credibility. A survey based upon non-probability sampling methods may produce the same results as a probability survey, but the results will be harder to defend against skeptics who suspect that the findings may reflect poor sampling rather than actual behavior. Probability methods produce data that can be interpreted with much greater confidence. This should in turn translate into a firmer basis for decision-making in designing HIV prevention programs and in allocating resources.

Early in the HIV epidemic, much research was conducted on an ad hoc basis – a response to the need for ANY information, as quickly as possible. More recently, however, the demand has grown for the systematic collection of high-quality data that can be interpreted and acted upon with greater confidence. This demand has spurred the development of methods to extend probability sampling as much as possible to surveys of population sub-groups that are difficult to enumerate. It is acknowledged, however, that the use of probability sampling methods will not be feasible for some populations: notably, those whose members do not congregate in fixed locations and for whom it is thus not feasible to develop a list of sampling frame. When a sampling frame cannot be constructed, the use of non-probability sampling methods is the only alternative.

End of excerpt

As describe in the excerpt above, random (or probability) sampling is the only method that will yield survey results that:

- Are systematic and replicable
- Can produce estimates with minimum bias
- Can measure trends over time

For this reason, only probability sampling method for BSS will be discussed in this chapter. If you choose to use a non-systematic method for conducting your survey, you must recognize in advance the problems inherent in employing a non-probability sampling methodology. These problems include: 1) difficulty in defining what population was captured in the survey and generalizing the results to the total population, 2) inability to calculate confidence intervals or measure trends over time, and 3) difficulty in comparing your results to other surveys.

The following is a description of the different types of probability sampling that can be used for populations sampled through households and how to decide what type of sampling methodology would be most feasible for your survey. Detail on how to operationalize each sampling method is provided later in the chapter.

A **simple random sample** is the 'gold standard' of sampling frames because it ensures that every household has the same probability of being selected for the survey. Using simple random sampling also assists your entire survey process in the following ways:

1. It does not necessitate inflating the sample size to account for the design effect caused by clustering (*for more information on design effect, see Chapter 11: Data analysis*)

2. It assists you in scheduling and supervising your field work because you know exactly how many and which housholds will be recruited each day, where the households are located, and how many interview teams will be needed in each area to complete the scheduled households.
3. It minimizes the number of decisions the team leaders will have to make about sampling when they are in the field, thus minimizing biases caused through sampling errors.
4. It allows the lead investigator and central supervisors to more easily document and monitor the sampling process.
5. It reduces the complexity of the survey analysis

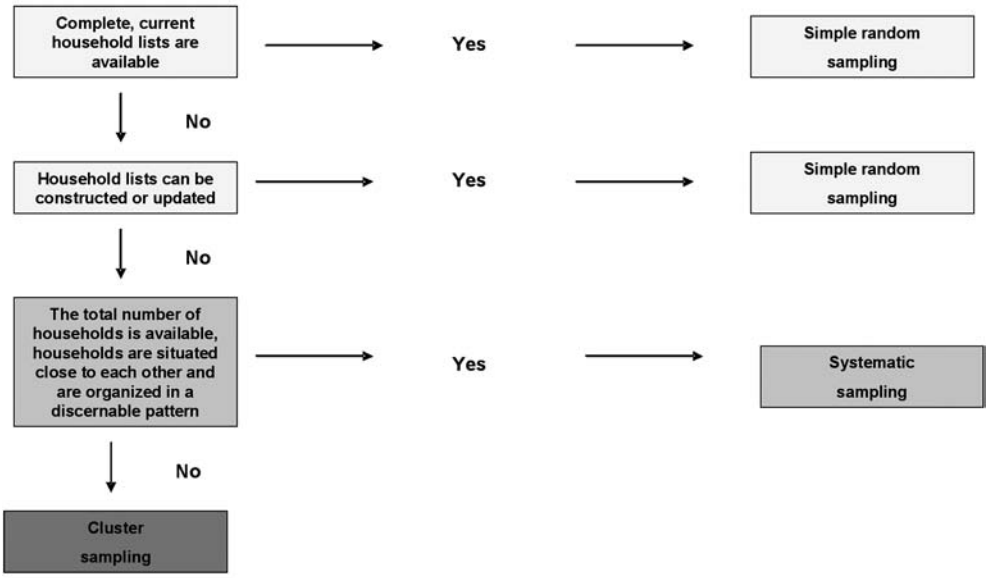
It is recommended that when household lists are not available, the populations are concentrated in a small, well-defined area, easily accessible areas, and if it is feasible to do so, that you construct your own list of households in the survey area or update already existing lists. This information is beneficial to a wide-range of organizations, and you should consider finding partners in the area who would assist you or undertake the project themselves.

If household lists are not available, but the total number of households in the population is available and the households are well-organized and close together (for example, tents in refugee camps), then **systematic sampling** can be used. Systematic sampling allows you to determine, based on the total number of households in the community and the number of households that need to be selected for the survey, that after beginning from randomly selected household Z, every Xth household after that should be recruited for the survey. More information on how this process is done and whether it is viable for your study area is discussed later in this chapter.

Systematic sampling has many of the same benefits of simple random sampling, including smaller sample sizes than cluster sampling and a less complex analysis. However, it does not enable you to construct a sampling frame prior to the beginning of data collection so it is more difficult to schedule and supervise a systematic sample then a simple random sample. Hence it is more open to 'improvising' in the field when survey teams encounter unexpected circumstances.

While simple random sampling is the ideal and systematic sampling is second in-line, most often the survey populations are spread out over a large area and it would be too time consuming and resource intensive to create a complete household list or travel to households that are not grouped together. When this is the case, a **cluster sampling** methodology should be used. What this means is that you will randomly select groups of households instead of selecting individual households. This sampling method does not require you to have a complete list of every household that falls in your survey area or that the households are close together organized in a discernable pattern.

Figure 2: Choosing a sampling methodology



The following section provides descriptions of sampling methods and is **excerpted from** Measuring, US Centers for Disease Control and Prevention and the World Food Programme, 2005. http://www.unhcr.org/publ/PUBL/45f6abc92.pdf

Excerpt

Simple random sampling

Simple random sampling is the process in which each sampled unit (person or household) is selected one at a time from a list of all eligible sampling units. Simple random sampling can be done many ways, but it requires a complete list of all the basic sampling units in the population.

The prototypical method for simple random sampling is to write the name of each household or person on a piece of paper, then put all the pieces of paper in a bowl and randomly select as many pieces of paper as the number of households or people needed for the survey. Of course, if there are 10,000 households in the population, it would take a very long time to write the names of the households on 10,000 pieces of paper.

Another way would be to number the households. Then random numbers between 1 and the total number of basic sampling units could be selected from a random number table or generated on a calculator or computer. The household with the same number as the random number would be included in the sample. The quantity of random numbers selected equals the desired number of sampling units, as determined by the sample size calculation.

Systematic random sampling

Systematic random sampling is an alternate way to select a random sample. It may be much easier when the population is large. Systematic random sampling selects one household at random, then selects every nth household thereafter, where "n" equals the sampling interval. The sampling interval is the total number of sampling units in the population divided by the desired sample size. The first household is selected by choosing a random number between 1 and the sampling interval. Then the sampling interval is added to that number to select the next sampling unit. For all subsequent selections, the sampling interval is added until the end of the list of sampling units is reached.

For example, if there are 10,000 households in the population and a sample of 1,000 households is needed, the sampling interval is 10 (10,000 divided by 1,000). A random number between 1 and 10, the sampling interval in this example, is chosen to select the first household. Let us say that 3 was the random number. The second household is selected by adding the sampling interval of 10 to the random start of 3 to choose household number 13. Then 10 is added to this number to select household number 23, and so on.

Systematic random sampling can be done with a list of all sampling units either on paper or on a computer. You can also arrange the sampling units, such as households, in a pattern so that they can be easily counted on the ground.

For example, in a refugee camp, the desired sample size was 350 households. The camp contained 5,263 households; therefore, the survey team needed to include in the survey sample every 15th household (5,263÷350=15). The team chose a random number between 1 and 15 inclusive to select the first tent, then beginning in one corner of the camp, counted tents along the first row until that number was reached. This was the first tent selected for the survey. Then proceeding down that row and subsequent rows of tents, the survey team counted each tent and selected every 15th tent to include in the survey. The team did not have a list of all households in the camp, but because the tents were in rows that were easily counted, a systematic random sample could be selected easily.

However, as mentioned above, in many populations, no list of all people or all households exists, and the dwellings are not arranged in neat rows. In such populations, survey workers may need to do cluster sampling.

Cluster sampling

Cluster sampling is used in large geographically dispersed populations, where no accurate list of households is available for the entire population and households cannot be visited systematically. This is the most common situation in most populations. Cluster sampling is often more convenient and uses fewer resources for trans-

port than simple random sampling because a cluster design reduces the distance the survey team has to travel between households. However, the sample size is usually larger than simple random sampling so that more households need to be visited.

With cluster sampling, the sampling is split into multiple stages. In most cluster surveys, there are two stages of sampling:

1. The first stage of sampling selects collections of persons or households, such as geographic areas within the population to be surveyed. These geographic areas may consist of political subdivisions, such as districts, sub-districts, census blocks or other defined areas. Alternately, in rural regions, the areas to be selected may be discrete concentrations of population, such as villages.

2. The second stage of sampling chooses, within each selected area, the households to be included in the survey.

Stage one: selecting the primary sampling unit

Cluster sampling requires the division of the population into smaller geographical units, such as city blocks, sub-districts or villages. These geographic units may be referred to as enumeration areas, enumeration units, or other terms, and are usually available from the organization within the country responsible for the census.

A sample of these geographic units, called primary sampling units (PSUs), is then selected in the first sampling stage. Although many texts and people use the word "cluster" to mean the same thing as PSU, they are not the same. The PSU is the geographic unit selected during the first stage of sampling. A cluster is the group of basic sampling units, usually households, chosen within a selected PSU. Therefore, when carrying out the first stage of sampling, you are selecting those PSUs in which a cluster will be selected; you are not directly selecting clusters.

Probability Proportional to Size (PPS)
When determining what to use as PSUs, the selection of smaller PSUs rather than larger PSUs will facilitate second stage sampling. However, in order to use a geographic unit as PSUs, some measure of the relative size of each unit, most commonly the population or the number of households, must be known.

For example, if a survey is done in a city, it would be better to use city blocks rather than larger neighborhoods as PSUs. In addition, each geographic unit should have at least the number of households required to form a complete cluster. Moreover, the geographic unit chosen to be PSUs must be mutually exclusive; that is, all households in the population being surveyed must be contained in one PSU and only one PSU.

The first stage selection of PSUs is done so that the chance of any specific PSU being

selected is proportional to the size of that PSU relative to the entire population. This type of selection is called "probability proportional to size" or PPS. Thus, if one PSU has a population of 5,000 and another 1,000, then the former PSU has five times the chance of being chosen to contain a cluster as the latter PSU. This is the main reason why some measure of the size of each PSU is required.

PPS sampling is demonstrated schematically in Figure 3. Each row of boxes represents 6 PSUs from which we must select one randomly. Each PSU, A through F, is assigned to a box, and the population of that PSU is written inside that box. In the first row, the size of the box allocated to each PSU is equal. As a result, if we throw a dart randomly at the first row of boxes, the chance that PSU A will be hit is the same as the chance that PSUs B through F will be hit. As a result, each PSU has the same chance of selection as any other PSU. This is called selecting PSUs with equal probability.

On the other hand, if we make the size of the box for each PSU proportional to the population of that PSU, as is done in the second row of boxes, the chances that a randomly thrown dart will hit PSU A is much smaller than the chance that it will hit PSU C, which has a much larger population. This results in PPS sampling; that is, the probability of selection for any single PSU is directly proportional to the size of that PSU relative to the sizes of all other PSUs.

Figure 3: Schematic demonstration of 1st. equal probability, and 2nd. probability proportional to size

	A	B	C	D	E	F
Not PPS	231	912	3,099	376	484	763

	A	B	C	D	E	F
PPS	231	912	3,099	376	484	763

In practice, we cannot draw boxes for each of the PSUs from which the selection will be made in a real survey. Moreover, throwing a dart is not really a random way of making a selection. When carrying out sampling for a cluster survey, the following method is usually employed to select PSUs with probability proportional to size:

1. Determine the sample size using methods described above.

2. Obtain the best available data on subdivisions of the populations which might be used as PSUs. You will need the size of each unit as well as the location of each unit. Census data from the national or local government offices often provide such information. In stable populations with little in- or out-migration, a census that is several years old may still be acceptable. In emergency situations, data may consist of population estimates or registration data in camps. Alternatively, if no population data are available, you can estimate the relative size of the population living in each PSU using data from key informants such as community leaders or health workers.

3. Create a list of PSUs with a column containing the measure of the size of each PSU. The order in which PSUs are listed is not important. [Usually it is recommended to sort on a variable of interest for implicit sampling, such as on rural/urban status, or by geographic areas.] What is important is that all PSUs be included on the list. Figure 4 gives a hypothetical example of a list of villages with the number of households in each village.

4. Add columns to this list which contain the cumulative population. Imagine that you are numbering all the households in the entire population. In Figure 4, households number 1-600 are located in the village of Utural. Households number 601-1300 are located in the village of Mina, etc.

5. Determine the sampling interval. The sampling interval is the number of PSUs you will select divided by the total number of households in the population. For example, if you wanted to select 30 PSUs from the population listed in Figure 4, the sampling fraction would be 25,370 / 30 = 845.7.

6. Select a random number between 1 and the sampling fraction. The village where this number household is located will be the first PSU containing a cluster. For example, if 399 is the random number, household number 399 is located in Utural, and the first cluster will be located in that village.

7. Add the sampling fraction to the random number selected above. The village where this number household is located will be the second PSU containing a cluster. For example, 399 + 845 = 1,244. Household number 1,244 is located in Mina; the second cluster will be in this village.

8. Continue adding the sampling fraction to the previous number to determine where the remaining clusters will be located. Some PSUs may be large enough to be selected more than once to contain a cluster. In such a case, when the survey team arrives at this PSU, they should select as many clusters from this PSU as indicated during the first stage sampling procedure.

The selection of each cluster in a PSU with more than one cluster should be completely independent. The procedure for selecting households will be described below.

Figure 4: A list of hypothetical villages to be included in a cross-sectional survey

No.	Village name	Number of households	Cumulative number of house holds-Lower	Cumulative number of house holds-Upper	Cluster number
1	Utural	600	1	600	1
2	Mina	700	601	1,300	2
3	Bolama	350	1,301	1,650	
4	Talum	680	1,651	2,330	3
5	War-Yali	430	2,331	2,760	
6	Galey	220	2,761	2,980	4
7	Tarum	430	2,981	3,410	
8	Hamtato	150	3,411	3,560	
9	Nayjaff	90	3,561	3,650	
10	Nuviya	300	3,651	3,950	5
11	Cattical	430	3,951	4,380	
12	Paralai	150	4,381	4,530	
13	Egala-Kuru	380	4,531	4,910	6
14	Uwanarpol	310	4,911	5,220	
15	Hilandia	2,000	5,221	7,220	7, 8, 9
16	Assosa	750	7,221	7,970	
17	Dimma	250	7,971	8,220	10
18	Aisha	420	8,221	8,640	
19	Nam Yao	180	8,641	8,820	
20	Mae Jarim	300	8,821	9,120	11
21	Pua	100	9,121	9,220	
22	Gambela	710	9,221	9,930	12
23	Fugnido	190	9,931	10,120	
24	Degeh Bur	150	10,121	10,270	
25	Mezan	450	10,271	10,720	13
26	Ban Vinai	400	10,721	11,120	
27	Puratna	220	11,121	11,340	
28	Kegalani	140	11,341	11,480	14
29	Hamali-Ura	80	11,481	11,560	
30	Kameni	410	11,561	11,970	
31	Kiroya	280	11,971	12,250	15
32	Yamwela	330	12,251	12,580	
33	Bagvi	440	12,581	13,020	
34	Atota	320	13,021	13,340	16
35	Kogouva	120	13,341	13,460	
36	Ahekpa	60	13,461	13,520	
37	Yondot	320	13,521	13,840	
38	Mozop	1780	13,841	15,620	17,18,19
39	Mapazko	390	15,621	16,010	
40	Latohah	1,500	16,011	17,510	20,21
41	Voattigan	960	17,511	18,470	22
42	Plitok	420	18,471	18,890	
43	Dopoltan	270	18,891	19,160	23
44	Cococopa	3,500	19,161	22,660	24,25,26,27
45	Famegzi	400	22,661	23,060	
46	Jigpelay	210	23,061	23,270	28
47	Mewoah	50	23,271	23,320	
48	Odigla	350	23,321	23,670	
49	Sanbati	1,440	23,671	25,110	29,30
50	Andidwa	260	25,111	25,370	

How do you decide how many clusters should be selected?

For a given sample size, the more clusters in the survey sample, the lower the design effect and the greater the precision obtained by the survey. However, a larger number of clusters often requires substantial additional logistic and transport costs because travel between clusters may be difficult or involve long distances. Determining the

number of clusters to use in a survey therefore requires weighing the advantage of greater precision with the disadvantage of greater cost.

Studies have demonstrated that, for a given overall sample size, having fewer than 25-30 clusters may lead to a high design effect and an unacceptable loss of precision. Adding more than 30 clusters often does not increase the precision enough to justify the additional cost. Consequently, many surveys use 30 clusters.

The size of clusters is determined from the number of clusters, as discussed above, and the sample size calculation. The total sample size is divided by the number of clusters to decide how many basic sampling units should be included in each cluster. For example,
if sample size calculations determined that 450 households should be selected for a survey and 30 clusters were to be used, then each cluster would contain 15 households.

Stage two: selection of households to form the clusters

There are several methods of choosing the households within the selected PSUs.

1. Simple or systematic random sampling

The best way is to select households using simple or systematic random sampling, as described previously. Simple random sampling can be done using a complete list of all the households in the selected PSU. Village leaders sometimes have such a list available to keep track of tax obligations or for some other purpose.

If there is no written list, village leaders or elders can often tell survey workers the names of the heads of all the households in the village while survey workers write them down. When creating such a list, survey workers must be very careful to ensure that the informants have not forgotten any household in the PSU, such as households headed by women, households of poor people, households of ethnic minorities or others. Once a list is located or created, households can be numbered and random numbers used to select individual households for inclusion in the survey.

Another method of random sampling within PSUs is to select households from the household list using systematic random sampling. Alternately, survey workers can draw a rough map of all the households within the PSU and then carry out systematic random sampling using the map to ensure that no households are missed by the sampling.

2. Segmentation

If the PSU is large enough to make the techniques above too time-consuming, the PSU can be divided into segments of roughly the same size. One of these segments

is then chosen at random. In general, the segments should contain fewer than 250 households. These households then are listed and the required number of households is selected.

3. EPI method

If it is absolutely not possible to select the households using random or systematic sampling, then the sampling method frequently used by WHO's Expanded Programme on Immunization (EPI) can be used. While this method is simple, widely known and rapid, it may result in a biased sample. To employ this method, the following procedures should be followed after arrival at the selected PSU:

1. Go to the center of the selected PSU. Local residents can help you locate the center.

2. Randomly choose a direction by spinning a bottle, pencil or pen on the ground and noting the direction in which it points when it stops.

3. Walk in the direction indicated by the pen, from the center to the edge of the village, counting the number of houses on the way.

4. Draw a random number between 1 and the number of houses counted on the line. Count the number of houses on the line from the center of the village; the house matching the randomly selected number is the first house to be included in the survey.

5. Interview all eligible people 15-49 years who consent to take part in the survey

6. Subsequent households are chosen by proximity. In a village where the houses are closely packed together, choose the next house on the right. If the village is less densely inhabited, choose the house with the door closest to the last house surveyed, whether it is on the right or left.

Regardless of the method of selecting households, if there are two clusters located in the same PSU, their selection should be completely independent. For example, if you make a list of all households in a village, you should select one cluster by random sampling, then select the households for the second cluster. The lists of households for each cluster should be kept separate.

Some potential operational problems

Implementation problems can arise in even the best-planned surveys. Typical among these are inaccessible clusters, non-response and an insufficient number of households in a given PSU to complete an entire cluster.

Inaccessible clusters

At times, it may be impossible to reach a sample cluster due to poor weather, impassable roads, insecurity or other reasons. Usually, the best recourse is to replace the cluster with another randomly chosen cluster with similar characteristics. For example, if the cluster in question is located in the far northern part of the area included in the survey, it should be replaced with another cluster in the same general area, but one that can be reached during the period of survey fieldwork.

To minimize the risk of bias, replacement clusters should be chosen from among similar clusters; convenience should not be an issue. As far as possible, supervisory personnel should make decision on replacement clusters.

Survey non-response

Non-response is an important issue in surveys. When households are selected, there
may be non-response at two levels: (1) entire households may be missing or refuse participation, and (2) individuals within consenting households may refuse participation or be absent. The initial calculation of sample size should compensate for the predicted level of both types of non-response. This will help ensure that the final survey sample will have the required precision in spite of some non-response.

When no one is at home in a selected household, the survey team should inquire from neighbors whether the dwelling unit is inhabited and if so, where the residents are and when will they return. If they are to return before the survey team must leave the PSU, a message can be left that the survey team will return at a prearranged time. If the house is not occupied, no further action is required. No pressure should ever be applied to an individual within a selected household who refuses to participate.

Non-response can bias the survey results because people who participate in a survey may be systematically different than those who do not. These differences may be reflected in the indicators that are being measured. As a result, non-response should be minimized as much as possible by allowing adequate time to reattempt contact with absent household members. Moreover, the reasons for both household and individual non-response should be noted on data collection forms to allow the assessment of potential non-response bias during later data analysis.

Insufficient number of households

Before selecting the sample of PSUs, survey managers should go through the list of PSUs to be sure each is large enough to select a complete cluster. If certain PSUs are not large enough, they can be combined with adjacent PSUs. If this cannot be done before sampling because the list is too long, another procedure can be followed in the field to complete a cluster. Once all available households are selected in a small PSU, selection of households can be continued in a neighboring PSU to complete the cluster.

End of excerpt

Determining the number of households in each PSU

Because every eligible member of the household is going to be selected, you must determine prior to the start of data collection how many households will be selected in each PSU. In order to do this, you need to know the approximate household size for each of your populations and the approximate number of eligible participants (men and women aged 15-49 years) in the average household.

In most areas, organizations that provide services for the communities of interest such as UNHCR and the World Food Programme and government offices such as the Ministry of Health, will have estimates of this information. Gather as much local knowledge as is available through these organizations and from reports and other surveys done in the area. Make every attempt not to use household size information from different communities or areas of the country or to just to pick a number at random.

If your estimated household size (for eligible population members) is too small, you will end up sampling more households then was necessary. Consequently, your sample will be larger then anticipated and you may run over budget and over time. If your estimated household size is too large, then you may complete your entire sample and fall under the necessary sample size which will require you to resample. Either circumstance is undesirable, so it is worth collecting the local information ahead of time to make sure your estimated household size is as precise as possible.

Example of determining the number of households in each PSU

Based on locally available data, in one BSS site the estimated number of eligible participants (men and women aged 15-49) in the refugee households was 2.0. In the surrounding community the estimated number of eligible participants per household was 2.3.

The sample size of the survey was 800 for refugees and 800 for the surrounding community.

800/2.0 = **400 refugee households** to be selected.

800/2.3 = 347.8 or **350 national households** to be selected.

For additional information on other sampling methods, please refer to the sampling chapters in

Measuring and Interpreting Malnutrition and Mortality: A Manual, US Centers for Disease Control and Prevention and the World Food Programme, 2005.
http://www.unhcr.org/publ/PUBL/45f6abc92.pdf

Behavioral Surveillance Surveys: Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV, Family Health International, 2000
http://www.fhi.org/en/HIVAIDS/pub/guide/bssguidelines.htm

Deitchler M, Valadez JJ, Egge K, Fernandez S, Hennigan M. *A field test of three LQAS designs to assess the prevalence of acute malnutrition*. Int J Epidemiol. 2007 Aug;36(4):858-64. Epub 2007 May 21.

Reinhard Kaiser, Paul B. Spiegel, Alden K. Henderson, Michael L. Gerber (2003) *The Application of Geographic Information Systems and Global Positioning Systems in Humanitarian Emergencies: Lessons Learned, Programme Implications and Future Research*; Disasters 27 (2) , 127–140 doi:10.1111/1467-7717.00224

Checchi F, Roberts L :*Interpreting and using mortality data in humanitarian emergencies: a primer for non-epidemiologists.* HPN Network Paper 52, Overseas Development Institute 2005. http://www.odihpn.org/report.asp?ID=2749

Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations: The SMART Protocol, January 2005, http://www.smartindicators.org/SMART_Protocol_01-27-05.pdf

Emergency nutrition assessment: Guidelines for field workers, Save the Children UK, 2004 http://www.savethechildren.org.uk/en/54_2320.htm

Medecins San Frontieres. Refugee Health: An Approach to Emergency Situations. 1997, MacMillan Educational Ltd.,. London

Checklist 2: Sampling

- ☐ Define your sampling universe
- ☐ Determine the key indicators that you want to measure, with what precision you want to measure them, and if you will be evaluating changes in these behaviors over multiple survey rounds
- ☐ Determine what type of sampling methodology you will use and whether you will have to increase your sample size to account for design effect.
- ☐ Calculate the sample size
- ☐ Choose the sampling methodology

Chapter 6

Questionnaire and informed consent

Designing appropriate questionnaires is a difficult task. There is always a struggle between wanting to use the opportunity to gain as much knowledge of the population as possible, but at the same time wanting to keep the questionnaires short and concise to avoid putting too much strain on the resources of the field team and the good humor of the participants. The questionnaires in this guide were refined by a team of program managers, epidemiologists, and BSS supervisors with years of combined experience in working with displaced populations and conducting behavioral research. Only questions which were found to be of direct value to program implementers were kept in the questionnaire. However, though the questionnaires have been carefully reviewed, potential problems may still exist.

Questionnaire structure

Two standardized questionnaires are provided in the appendices. The first is the **core questionnaire** (*Appendix 3*), composed of a series of questions on the following:

1. Socio-demographic characteristics
2. Alcohol and drug use
3. Circumcision
4. Sexual activity
5. Regular sex partners
6. Casual sex partners
7. Transactional sex partners
8. Forced sex
9. Anal sex
10. Male and female condoms
11. Sexually transmitted infections
12. Knowledge, opinions and attitudes towards HIV/AIDS
13. Exposure and access to interventions

The second is the **expanded questionnaire** (*Appendix 4*), which incorporates additional questions to the core questionnaire including the following:

14. Additional questions on circumcision
15. Military activity
16. Additional questions on transactional sex
17. Additional questions on sexual and gender-based violence
18. Additional questions on female condoms
19. Additional questions on HIV knowledge

The core questionnaire was designed to measure the key indicators of HIV risk and knowledge as well as provide the data necessary for developing interventions in response to the findings. The expanded questionnaire has selected optional questions which may be useful in many situations but not crucial to all. The additional questions in the expanded questionnaire are highlighted in blue.

❧ *For additional information on indicators, please see Chapter 2, BSS Objectives and Core Indicators*

Why are there two different questionnaires?

An example: While encouraging the use of male condoms has been a frontline tactic of controlling HIV transmission since the early stages of the epidemic, expanding access to female condoms is only now being recognized as another important risk reduction option.

In order to understand the level of knowledge about female condoms in the population, the core questionnaire asks: *Have you ever heard of a female condom?* In a many areas a majority of the responses to this question will be 'no' because female condoms have not been included in local interventions or mass media awareness.

As discussed earlier, the survey team should have a working knowledge of local program components and should be able to determine whether it is worth expanded upon this line of questioning. If female condoms have not been introduced at all in the community, then the proportion of respondents who have ever heard of one will likely be quite low. This will provide valuable baseline information for phasing in female condoms as a program component, but it will not yield enough positive responses to ask more detailed questions about this subject. However, if increasing access to female condoms is part of an intervention strategy, then the expanded section should be included in the survey. These questions in the expanded questionnaire ask: *Have you ever used a female condom? Would you/your partner be willing to use a female condom if available? Do you know where you can obtain a female condom?*

Behavioral surveys always need to take into consideration the local context in order to make them as useful as possible. There may be questions that are not in the questionnaire that are necessary in understanding the dynamics of HIV transmission and the program response in your locality. Some of these questions might include more detailed information about exposure to interventions, movement patterns, or interactions between the refugee and surrounding communities. If you do need to add questions to the questionnaire or alter it in any way, please carefully follow the guidelines in the box *below*.

Altering and revising questionnaires for population-context

- Do not add questions simply because you think the answers will be interesting. Consider how you and others will use the information. Will the responses bring added value to how you will respond to the HIV epidemic in your community?
- Whenever possible, do *not* remove existing questions or change time reference periods. Many of the questions have been designed to allow comparison across regions and populations, one of the core purposes of the surveys, and to conform to international reporting requirements. *For more information on time reference periods, see the learning point below.*
- Do not expect too much from the quantitative questionnaires. These measurement tools are not overly sensitive. They can only explain what happened and when it happened. They cannot clearly answer why someone was motivated to act in the way he or she did with the detail that we often want. Do not forget the important role of qualitative research in answering many of your questions.
- Avoid multiple-response, unprompted and open-ended questions whenever possible. They can be difficult to ask, difficult to code, and difficult to analyze.
- If you change the questionnaire, you will likely change the skip patterns. Review the revised skip patterns **carefully** before you finalize your tool.
- Very general questions often elicit information that is hard to interpret. Always be as explicit as possible about language and time frame. Read the questions to colleagues and ask how they would explain it. If your colleagues interpret it in different ways, the participants will certainly do the same.
- Do not try to ask for too much information in one question. Split the question into two or more parts if it helps to make it clearer. In the end this will save time because the responses will not be open to multiple interpretations.

Fieldwork forms

On the first three pages of the questionnaire, you will find sheets for **identification** information on the household and participant, the **interviewer's name and code**, recording **quality control** steps during the fieldwork, central level and during data entry, and **household and participants consent or refusal**.

The *household and participant identification information* is very important for the data analyst. It will be entered into the data set and used extensively, so it must be entered accurately.

The first information that should be completed is the questionnaire and household serial numbers. Every questionnaire in your survey should be given a unique, usually continuous, number. This number will be used during data analysis as the identifier for each individual participant, so there should be no repeated numbers. There is no one best way to determine

how the questionnaires number is assigned because it will often be determined by how the teams are supervised, how many teams are working on the survey, and how the data collection is organized. Sometimes the numbers are assigned to the questionnaires by the lead investigator before the teams leave for the field each day. Other times each team leader will have his or her own list of numbers to enter on the questionnaires. Or the survey investigator may decide on a better method for assigning and tracking questionnaire numbers. The most important aspect of the questionnaire number is that each one is unique for each participant. If the numbers are assigned in a continuous manner, this may help in data collection supervision and data entry management, but it is not required.

Each household in your survey should also be assigned a unique household serial number. This information will allow analysts to determine how many households were selected, how many eligible people there were in each household, and which participants in the data set belonged to which household. It will also help survey managers track data collection progress. If you select households from a complete list before data collection begins, then the household serial number can be assigned by one person, preferably the lead investigator, before the team goes into the field. However, if your survey is employing cluster sampling, then you will not be able to identify the households to be sampled ahead of time, and the field supervisors who are managing the sampling process will have to assign the household serial numbers. It is important that the lead investigator determines ahead of time how household numbers will be assigned and that the field supervisors clearly understand the process.

Questionnaire serial number	|___|___|___|___|
Household serial number	|___|___|___|___|

A numeric code should be determined for the country, regions/provinces that fall into the survey area, and each of the areas or villages of the camp and surrounding communities selected for the survey. All of this information must be entered into the data set as a *number*, not as a name (see the box below for additional information). The most efficient way to do this is for the lead investigator to draw up a list of all possible options for each identifier and assign a number to each option. If you are using other types of information to identify and select households for the survey, this should also be added to the questionnaire. The list then needs to be widely distributed to the field supervisors, interviewers, data analysts, and others using the information, and the field team must be trained on how to use it.

If you are conducting a cluster sample, the numeric code for the cluster in which each participant was selected should also be recorded. This information is necessary in order to adjust for the design effect introduced when doing a cluster survey (see *Chapter* 5 for additional information)

You may also want to identify whether a study site is considered ‘urban’ or ‘rural’. How these areas are defined will depend on your particular survey and geographic area, as well as considering standards used by government organizations such as the census bureau. Often capital cities and regional hubs are coded as urban and towns and villages are coded as rural.

It should be the team supervisor's responsibility to fill out all identifying information before giving the questionnaires to the interviewers. These identifiers are *critical* for survey analysis, and their correct completion should be managed very closely. Please note that this information is regarding where the interview is taking place; it is not information about the participant's home country/region.

IDENTIFICATION	
COUNTRY	|___|
REGION/ PROVINCE	|___|___|
CAMP/ SURROUNDING AREA (Camp = 1, Surrounding area = 2)	|___|
IDENTIFICATION NUMBER OF CAMP/ VILLAGE	|___|___|
CLUSTER NUMBER	|___|___|
URBAN/ RURAL (Urban = 1, Rural = 2)	|___|

The *name and code of the interviewer* should be recorded on the form, and it too will be entered into the data set. One of the most problematic areas in implementing behavioral surveys is interviewers who are either poorly trained or who do not clearly understand how to complete the questionnaires. When recurring mistakes in the way answers were recorded are identified during the analysis, it is sometimes possible to trace them to individual interviewers. Ideally, these mistakes should have been identified and rectified well before the analysis stage. However, this does not always happen. The ability for the study supervisors to track systematic inaccuracies back to a single study team member can also provide a strong incentive for increased rigor by team members.

NAME AND CODE OF INTERVIEWER

_____ |____|____|

The next section on the first page of the questionnaire is the *quality control* checks. The supervisor who is responsible for reviewing the questionnaire in the field should sign off and date the document when she or he has finished. When questionnaires are transferred from the study team to a central supervisor, again they should be reviewed, signed and dated. The data entry clerks will then do they same after they have completed entering the information from the questionnaire into the data base.

CONTROL				
	CONTROL ON FIELD LEVEL	CONTROL IN CENTRAL OFFICE	DATA ENTRY CLERK 1	DATA ENTRY CLERK 2
NAME DATE	----------|__| ----------	----------|__| ----------	----------|__| ----------	----------|__| ----------
REMARKS				

The final information to be recorded on the first page of the questionnaire is the date of the interview, as well as the starting and ending time of the interview (see below).

Date of interview: __/__/ day /__/__/month

Start of interview: __/__/ h /__/__/min

End of the interview: __/__/ h /__/__/ min

The second page of the questionnaire records information about recruitment of the household and the third page records information on the recruitment of the individual participants. This data is very important for monitoring and calculating non-participation rates and will be used by both the survey supervisors and the data analyst.

Before any participants can be recruited for the survey, the team leader will need to obtain consent from the head of the household. The team leader should record on the questionnaire of every person in the survey whether they are the head of household. They should also record the number of people living in the household and the number of people in the household who are aged 15-49 and eligible to participate in the survey.

The team leader should then record whether or not consent was given for household participation only for those people who are the heads of household. The team leader should also record if the house was ineligible because it did not fit the inclusion criteria, if the household was abandoned, if the household was temporarily absent, and if the household was entirely absent because it had moved, was repatriated, or members of the home were on extended travel. If the head of household refuses, this information should be recorded and that household will be considered completed.

HEAD OF HOUSEHOLD

PARTICIPANT IS HEAD OF HOUSEHOLD OR REPRESENTATIVE OF HEAD OF HOUSEHOLD
..... |___|

1 = Yes

2 = No

3 = No head of household or representative present

NUMBER OF PEOPLE IN HOUSEHOLD

Total number of people living in household ..... |___|___|
Total number of eligible people aged 15-49 living in household ..... |___|___|

RESULT OF HOUSEHOLD RECRUITMENT

RESULT OF HOUSEHOLD RECRUITMENT..... |___|
1 = Head of household agreed to household participation
2 = Head of household refused household participation

Reason for refusal ……

3 = Household not eligible
4 = Household temporarily absent
Date and time of first visit .....
Date and time of second visit .....
Date and time of third visit .....
Reason for household's absence .....
5 = Household abandoned
6 = Household on extended travel
7 = Other (specify) .....

After consent has been obtained from the head of household, the interviewer will obtain consent from the individual participant and record that information in the result of participant recruitment box. The participant's relationship to the head of household, sex and age should also be recorded. This information must be filled out for every eligible person recruited for the survey, whether they agreed to participate or not. The information on sex and age will help to determine whether there were more refusals or absences among certain sexes or age groups and if this may bias the survey results.

RELATIONSHIP TO HEAD OF HOUSEHOLD

PARTICIPANT'S RELATIONSHIP TO HEAD OF HOUSEHOLD
.....|___|
1 = Head of household
2 = Spouse
3 = Son/Daughter
4 = Father/Mother
5 = Brother/Sister
6 = Other relative
7 = Living in household but not a relative

RESULT OF PARTICIPANT RECRUITMENT

RESULT OF PARTICIPANT RECRUITMENT..... |___|
1 = Agreed to participate and interview was completed
2 =Agreed to participate but interview was not completed
3 = Refused to participate
Reason for refusal ……
4 = Absent
Date and time of first visit .....
Date and time of second visit .....
Date and time of third visit .....
Reason for participant's absence .....
5 = Other (specify) .....

If participant is not recruited because refuses (3), is absent (4) or for other reason, record age and sex of non-participant:

Record sex of the respondent	1 = Male 2 = Female	|___|
How old are you? **Record age in years**	Record number of years 99 = DON'T KNOW	|___|___|

Why do we record all information as a number?

A number (numeric variable) is much easier to analyze than a name (string variable) and is less subject to transcription errors. For example, suppose the Kakuma refugee camp in Kenya was assigned the number '2'. There is only one possible way to enter the number '2' into a data set. However, 'Kakuma' could be written by the interviewer in many different ways, including changes in spelling and capitalization. It could also be entered by the data clerks in a variety of ways. Each of these variations would appear in the data set as a unique answer, and it is difficult to combine them together into a single answer during data management. When this happens, you lose the much of your ability to analyze the information contained in the variable.

Learning point: What is a time reference period and why is it important?

Our surveys want to measure what behaviors may be putting populations at risk for HIV transmission as well as when those behaviors occurred. Asking a 40-year-old man who has been sexually active since the age of 16 whether he has exchanged sex for money, a gift or a favor may be a usefully screening question, but it provides little other useable information. What if the only the only time he ever had transactional sex was at his sexual debut twenty years ago? Is this a behavior that is currently putting him at-risk for HIV? Would we know if it was a past or recent behavior from the way the question was worded? Would we be able to distinguish his response from a man who bought sex yesterday?

A time reference period establishes when a behavior occurred. It is usually phrased as "in the past week" or "in the past 12 months" or as another amount of time that gives the behavior temporal boundaries.

Time reference periods need to be meaningful for the population you are surveying and behavior you are measuring. In a population where a behavior is common, such as selling sex among female sex workers, the time reference periods are often shorter. A woman who has several sex partners a night will not be able to recall exactly how many times she sold sex in the past year. But she could provide a closer estimation of her number of partners in the past day or week. It is also important to make sure that the period of time is long enough that it is not unduly affected by the season, day of the week on which the question was asked, or other factor that would limit your the ability to generalize the responses.

Conversely, in a population where a behavior is existent but less common, the time reference periods are often longer. When trying to measure what proportion of a male population buys sex, asking about the past week might not provide many positive replies. However, asking them about the past 12 months would provide a better indication of the more recent transactional sexual behavior.

Without time reference periods, you will not be able to:

- Understand when a high risk behavior occurred
- Determine comparability of behaviors across surveys, sites and populations
- Measure trends over time
- Report behaviors according to international standards

There is no fixed rule as to how long time reference periods should be; often they are determined by operational research or by learning from the experiences of other surveys in similar areas or populations. The time periods in these questionnaires are based on international standards, but they were also determined to be applicable to the populations in our surveys.

Informed consent

On the third page of the core and expanded questionnaires, you will find the **consent form**. This form is very important. It ensures that the participant is clearly informed about the following:

- The interviewer's name and affiliation
- The purpose of the survey
- How that person was selected to participate
- Confidentiality
- How the results will be used
- The benefits and drawbacks to participation
- How long it will take to administer the questionnaire

Before the interview can begin, a member of the study team must administer the informed consent form to each individual in a private space. The potential participant then gives verbal consent to take part in the survey or declines to be interviewed. The interviewer or other study team member signs the consent form if the person agrees. Please remember that the participant does not sign the form. If s/he did, then the questionnaire would no longer be confidential.

The consent form in the appendix will give you a guide to what information needs to be presented during the consent process. If your survey is being reviewed by an ethical review board then it may also have its own guidelines for what should be included in the informed consent form. This is the minimum amount of information that the participant will need to make an educated decision as to whether he or she wants to participate in the survey. You will need to carefully consider what additional information may be required or important for your particular survey. The interviewers should be trained on the consent process just as they are on administering the questionnaire.

For additional information on informed consent, please see Chapter 14, , Research Ethics

Translating and testing the questionnaire and consent form

The questionnaires should be translated into all the main language groups used by the survey participants. Whenever possible, participants should be interviewed in the language they speak most comfortably, though second languages can be used among populations that are truly bilingual. Interviewers should **never** translate the questionnaire into another language during the interviewing process. It will be important to carefully document all languages spoken in your survey area and what proportion of the study population speaks what language. This will not only impact the decisions on what languages the questionnaire must be translated into, but also how many copies of each questionnaire will be needed in the field for each language.

Accurate translation of quantitative surveys tools is always necessary, but it takes on particular importance in measuring sexual and drug-taking risk behaviors. The words describing different types of sexual partners, sexual acts, and types of drugs, as just a few examples, can vary widely from community-to-community and population-to-population. Sometimes idiomatic slang can more clearly define a behavior to a group than the more generally accepted word. In translating the questionnaire, both the language(s) of the participants must be considered as well as local terms that most clearly explain your question to the participant.

Translators must have a strong knowledge of both the languages they are working in (usually English and the local language). Before they begin their job, it is imperative that translators be trained on the objectives of the survey, the meaning of each question, and the definitions of the less common words and potentially confusing concepts. Translators are frequently hired because of their language skills, but often they do not work in public health and may not understand the nuances of the questions.

For example, one survey was conducted among refugees who communicated in five distinct languages. In one of the languages, the words for 'circumcision' and 'immunization' were the same, but because the translator did not understand the objective of the questions he did not know which concept he needed to explain. In another language used in the survey, there was no direct translation for 'anal sex', and local researchers had to work with the translators to both define the question and to find the best words to describe the behavior. In a third survey language, the direct translation for 'sex' was implicitly understood by the population to mean sexual intercourse only between a husband and wife, and 'sex' in any other context had to be translated differently. For these reasons, it is often useful to translate in a team that includes members of the survey population or people who are very familiar with cultural and geographic variations in the survey language(s) as well as the objectives of the questions. In addition, while it may be helpful to refer to the translations of similar surveys done in other regions, the translation process should always be done anew for your particular area and linguistic differences.

Translation should be done directly into the master questionnaire shell so the questionnaire format is the same for all language groups. Translators should not reformat the document because this creates confusion and can lead to missed questions and mistakes in skip patterns.

After a thorough translation has been done, the questionnaire should be field tested among the survey population. Field testing should be done *before* the questionnaire is finalized and *before* the interviewers are trained. Though this may seem obvious, often survey teams underestimate the amount of time it takes to translate and test the questionnaire. Always build in adequate time for this process. A correct translation is absolutely critical, and surveys done in multiple language groups make the entire process even more time-consuming. If the translation work is rushed and steps are missed, you may later find significant problems with the final tool that cannot be rectified once field work has begun.

Field testing should be done among people that represent a cross-section of the survey population. People representing different sexes, age groups, language and cultural groups, displaced populations and local community members, and other social factors that may affect the understanding of the participants should be included. Whenever possible the people on whom you test the questionnaire should not be part of the sampling frame. When this is not possible (for example, the entire community is in the sampling frame), they will no longer be eligible to participate in the survey in case they are selected during sampling.

The team conducting the field tests should administer the questionnaire and informed consent in the same manner they would during actual data collection. They should carefully record any inconsistencies in the tool and confusion that they or the participant had about any of the questions or skip patterns. They should then discuss the difficult areas with the participant and solicit his or her suggestions about what would make the questions clearer.

After testing, the questionnaires should be reviewed and all of the problems found in the field must be addressed. If significant changes are made, then it may be necessary to do an additional test of the revised field tool, or at least the problematic areas. Make sure there is enough time built into your schedule for survey preparation to account for these types of potentialities.

When the translated questionnaire has been satisfactorily revised, it should then be **back-translated** to its original language by someone other than the person who did the original translation in order to minimize bias. It is useful to have a person who is familiar with the terminology but not the actual survey tool to do this translation. The back-translated questionnaire should then be reviewed against the original tool to make sure that the meaning of each question and term was not altered during the translation process. This process is also important because the questionnaire may change during the pilot testing.

Make sure you set aside enough time for accurate translation, field testing, back-translation and printing of the questionnaires *before* you begin to train the survey teams. Do not underestimate how long this process may take and do not skip steps! A weak questionnaire will yield poor results.

Checklist 3: *Questionnaire and informed consent*

- ☐ Review core and expanded questionnaires
- ☐ Chose the best questionnaire for your location and needs
- ☐ Determine whether there are any other mandatory questions missing from the questionnaire
- ☐ Add new questions or otherwise revise the questionnaire if necessary
- ☐ Train the questionnaire translators
- ☐ Translate questionnaires and consent forms into all locally relevant languages
- ☐ Develop numeric codes for participant identification information
- ☐ Field test questionnaires and consent form
- ☐ Revise questionnaires and consent form based on results of the field test
- ☐ Back-translate questionnaires and consent form to the original language
- ☐ Compare the original and back-translated tools to ensure that no definitions or question meanings have changed
- ☐ Finalize and print questionnaire and consent form

Chapter 7

Survey documentation

There are several important parts to documenting your survey methodology and processes.

The first part is developing a survey **protocol.** Much of the information to be included in the protocol is already laid out for you in this manual, but it needs to be adapted for your local setting and based on the decisions you have made on specific survey objectives, sampling frame and sample size, what questionnaire to use, the definition of your population groups, what areas are included in your survey, and other key aspects of the survey methods. While all of this information should also be included in the final report, it should be documented *before* data collection begins.

People often think of protocols as large, time-consuming, technical documents, but they are in fact the vital framework for the survey. If you are not able to document the objectives, methods, fieldwork tools, ethical issues, and methods for analysis and data use, then you are not ready to be implementing the survey. Putting time in upfront to develop a detailed protocol will save you the headache of having to recreate your methods after data collection has been completed. The protocol will also give you a document that can be circulated for comments to people who have experience in behavioral data collection.

In addition to being a key step in preparing for data collection, the survey protocol is a necessary document to provide to ethical review boards (for more information see *Chapter 14: Research Ethics*), Ministries of Health, and other decision-making bodies. It will assist you in comparing your methodologies to those of the national surveillance system and other surveys to determine whether your data will be comparable. And it will also allow others besides the lead investigators to better understand how the data can be analyzed, interpreted and used.

Remember, one of the objectives of these surveys is to measure changes in behavior over time. In order to measure trends, the survey needs to be replicable. And in order for the survey to be replicated, the same population definitions, geographic boundaries, and methodologies from the baseline survey must be used. If this information is not well-documented then you will lose the ability to fulfill one of your most important objectives.

Key elements of the survey protocol include:

1. Names of investigators, affiliations and contact information
2. Background and rationale
3. Survey objectives
4. Definitions of survey populations and inclusion criteria
5. Geographic boundaries of survey
6. Sample size calculations and final sample sizes
7. Sampling methodology
8. Process for recruiting participants

9. Survey team composition
10. Training plans
11. Field monitoring and quality control checks
12. Informed consent
13. Questionnaire
14. Any services provided to participants such as health education or condoms and lubricants
15. Data management and analysis
16. Potential risks and benefits to participants
17. Ethical review and other ethical issues
18. Dissemination
19. Preliminary timeframe

In addition to developing a survey protocol prior to data collection, it is also necessary to develop an **operational manual**, or **field manual** that explains the step-by-step process of data collection, with a particular emphasis on the roles and responsibilities of each team member. Again, this does not need to be a lengthy, cumbersome document. The intention of the field manual is to explain what the survey teams must do in order to "operationalize" the protocol (who does what, when, and how?), and it will become one of the bases for your training of field staff. Teams should then have copies of all or parts of this information to reference when they are out in the field.

It is often easier to develop this reference not as a lengthy text document, but by creating a schematic of each step in the survey process, when that step happens, and who is responsible for it. Then include the supporting materials for each step. An example schematic of the steps in the data collection process is shown below. The other documents that would need to be included with in the field manual are the final sampling frame, the final household (or clusters) selected, the final data collection schedule, the final informed consent and questionnaire, household and participant information sheets, and guidelines for editing questionnaires. You should already have done most of this work while preparing for the survey and developing the protocol, so the field manual should be the last step in pulling together all of the information before beginning the training.

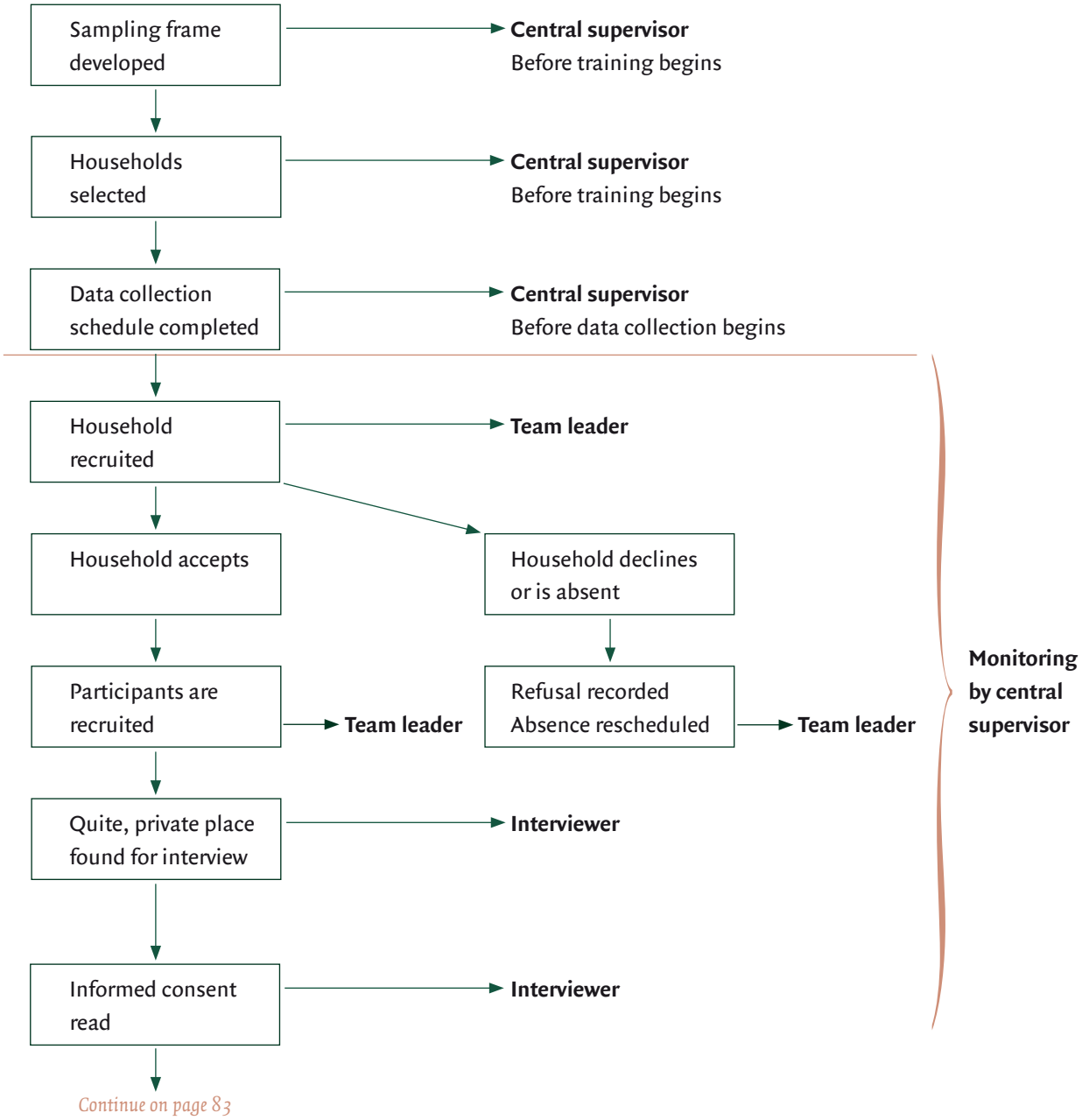
Why is it important to document your survey methodology and processes?

- People besides the primary investigators must be able to understand the survey methodology and analysis when interpreting and using the results
- The survey cannot be replicated if it is not documented and trends cannot be measured over time
- The data collection process must be standardized so the teams understand their responsibilities and do not have to improvise in the field
- Analysts may have to adjust for sampling design, and they cannot do so without the proper fieldwork forms
- Ethical review boards, Ministries of Health and other bodies will need to be able to assess the value and ethical processes of your survey

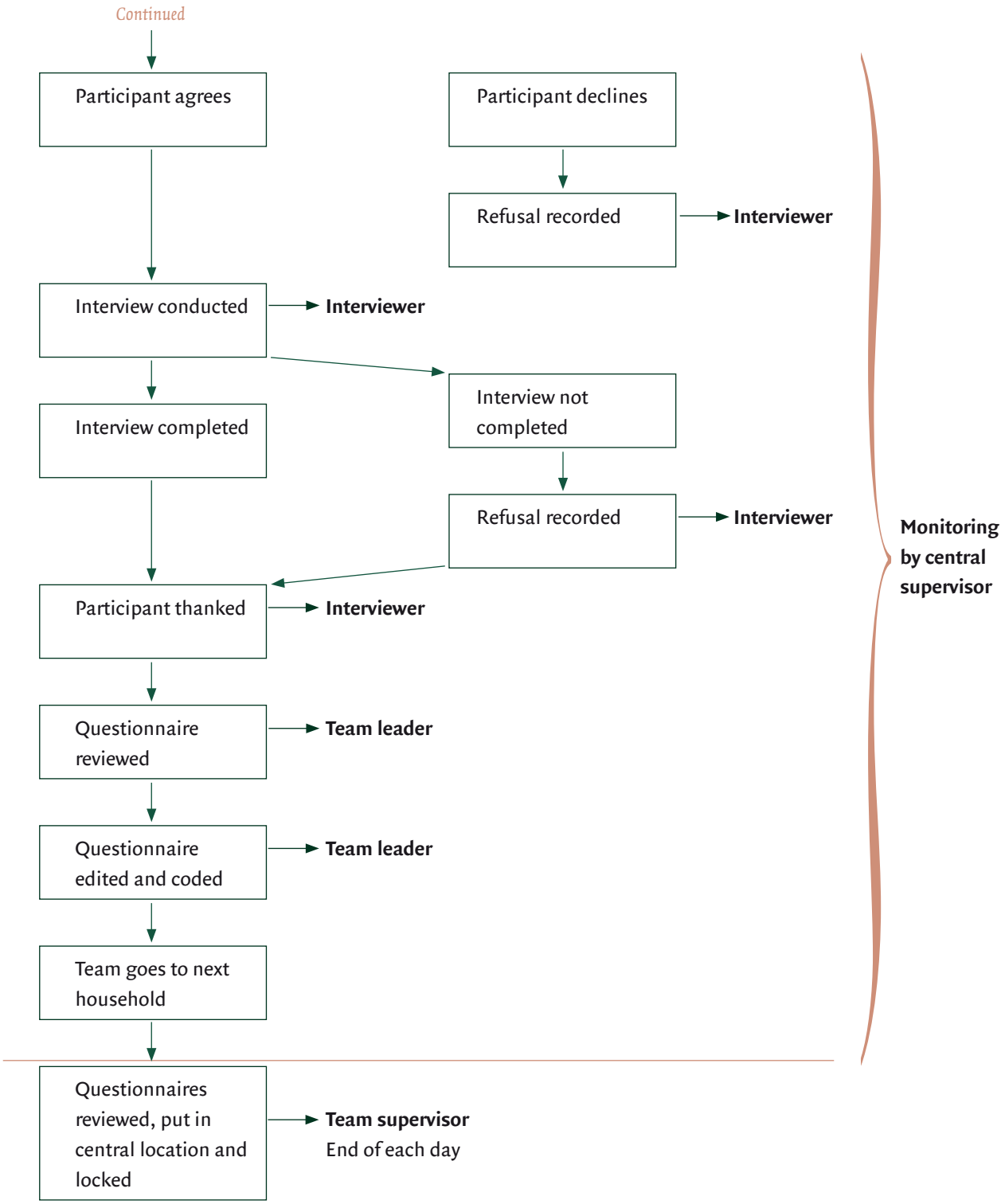
Checklist 4: Survey documentation

- ☐ Develop survey protocol
- ☐ Circulate protocol to other experienced researchers for feedback
- ☐ Document the steps of the survey, timelines, roles and responsibilities
- ☐ Develop field manual with all final documents

Figure 5: Example of data flow, responsibilities and time frame for data collection



Continue on page 83


Continued
Participant agrees
Participant declines
Refusal recorded
Interviewer
Interview conducted
Interviewer
Interview not completed
Interview completed
Refusal recorded
Interviewer
Participant thanked
Interviewer
Questionnaire reviewed
Team leader
Questionnaire edited and coded
Team leader
Team goes to next household
Questionnaires reviewed, put in central location and locked
Team supervisor
End of each day
Monitoring by central supervisor

Chapter 8

Field team composition and training

Field team composition

The number of people on the data collection team will be determined by the sample size, timeframe for data collection, distance between interview sites, and budget. As a rule of thumb, interviewers can be expected to do approximately five 30-45 minute interviews each day if they do not have to travel long distances between participants. There should be at least one supervisor for each of the field teams (referred to as the 'team leader' in this manual), and at least one central supervisor to assist all field teams. The central supervisor should be not be the lead investigator because that person will be offering oversight to the entire survey process. If you will be traveling by car, then drivers must also be considered. While this manual uses terms such as 'team leader' and 'central supervisor' to define what each person's position is, your team may want to use different titles or outline different roles for each team member based on the size and skills of your staff. The important thing is not what each person is called, but that they understand what their responsibilities are in the survey.

Below is an example of the members that should comprise the data collection team. Additional information on the entire survey team composition is provided in Step 8: Define survey team composition and members' roles and responsibilities in *Chapter 4: Survey preparation*.

- **Lead investigator:** There should be a single person in charge of technical oversight of the survey.

- **Central supervisors:** There should be at least one central supervisor for each of the survey populations (for example: one for refugees and one for surrounding communities). This person assists the lead investigator in managing the fieldwork, provides daily supervision to each of the field teams under his or her oversight, and liaises between the team leaders and the lead investigator. The central supervisors report directly to the lead investigator.

- **Team leaders:** Team leaders directly manage the interview teams in the field, and every data collection team should have one team leader. This person is responsible for managing the sampling, household consent, and checking the questionnaires, as well as directly supervising the interviewers on his or her team. However, the team leader should not conduct the actual interviews. The team leaders should report directly to the central supervisor and/or lead investigator.

- **Interviewers:** The number of interviewers per team may vary, but it is best not to exceed four interviewers per team leader. The interviewers administer participant informed consent and the questionnaire and report directly to the team leader.

- **Drivers:** Every team, or every group of teams working in the same geographic area, should have a dedicated car and driver for their work. The driver should remain with the team for the entire day.

In addition to the data collection staff, there will be data entry clerks and analysts on the team, though they will not necessarily be in the field during interviewing. Their roles are discussed in *Chapter 10: Data Management* and *Chapter 11: Data Analysis.*

Some important considerations in determining team composition include:

- Gender – In surveys of this nature, it is best to have women interviewing women and men interviewing men. Make sure this is true in your survey site when determining the number of interviewers of each gender you will require. If half of your participants will be women, then half of your interviewers should be women. However, the language and educational requirements necessary for the interviewers may make it difficult to recruit qualified female staff. If it is not possible to recruit enough women for your survey team, a female participant will need to be screened prior to the interview to assess if she would be willing to have a male interviewer. If she says no, then the interview must be postponed until a woman is available to conduct it. Whenever possible, it is best to recruit enough female interviewers prior to the training, but you should not compromise on the important skills of the interviewers to reach a gender quota.
- Languages spoken – You will likely be working in several different languages, and the languages spoken by the interviewers should proportionately represent those of the participants.
- Member of displaced or host population – As with gender, often participants feel more at ease with someone with whom they can identify. This may mean that it is best to have refugees interviewing refugees and the same for the members of the surrounding communities. However, this may not always be true, and sometimes participants would prefer to be interviewed by someone who is seen as totally outside of their communities and therefore less likely to compromise their confidentiality. Frequently, the number of displaced or host population interviewers will be predetermined by the languages used in the questionnaire. There is no fixed rule for who the best interviewers will be for your survey. It will be necessary to do pre-survey qualitative assessments to inform these decisions.

Affiliation – While many of the people at your site who have the necessary skills to be interviewers (see below) are often affiliated with NGOs or health centers, using such people can sometimes be problematic. If a participant knows that his or her interviewer is a health care worker or a member of an HIV prevention organization, then he or she may be more inclined to downplay his or her risk behavior, which leads to reporting bias. Whenever possible, use interviewers who will be considered neutral in their opinions to the participants. When NGO or health staff are used as interviewers, they should not be recognizable to the participant or report their affiliation.

Example: Determining the number and characteristics of field team members

The sample size in this example survey is 1,000 for the host population and 1,000 for the refugee population, of which half the sample are men and half are women. The host population speaks only one language, while there are three main languages in the refugee camp. It is expected that each interviewer will complete 5 interviews a day and there are 15 days allocated for data collection.

To calculate the number of interviewers for the host population, you know that all of the interviewers must speak and read Turkana. There will be 500 woman and 500 men in the survey. 500 women/15 days/ 5 interviews a day = 7 female interviewers for the host population. You would also need 7 male interviewers for this population. If these seven interviewers divide into two teams, then there should be at least one team leader per team and one central supervisor providing support to them.

To calculate the number of interviewers for the refugee population, you would need to do the same calculations, but you would also have to consider that the refugees speak three different languages. In this example, roughly half of the refugees speak Juba Arabic, about one quarter speak Dinka and the rest speak Somali. Find the total expected sample size for each language and conduct the calculations in the same way you did the host population. Each language group should have at least one team supervisor and there should be at least one central supervisor to support the teams interviewing the refugees. In this example, the survey team would be composed of 37 members.

Another important consideration in determining the time and staff that will be required for the survey will be revisits to households and participants who were absent during the time of recruitment. If the populations are situated close together, the selected households are easily identifiable, and you have only one or two languages in which to interview, then you may want to add additional interviewing teams whose primary task it is to revisit households and participants who were absent. However, if households are spread far apart, making travel and identification difficult, or there are numerous survey languages, you will want to build additional days into your data collection schedule for revisits to absent households and participants.

Table 3: Example calculations for the number of field team members by language and gender

	Host population	Refugee population			Total
Total sample	N=1,000	N=1,000			N=2,000
Language	Turkana (100%) N=1,000	Juba Arabic (45%) N=450	Dinka (23%) N=230	Somali (32%) N=320	
Number of interviewers					
- Female	7	4	2	2	15
- Male	7	4	2	2	15
Number of team supervisors	2	1	1	1	5
Number of central supervisors	1	1			2
Total field staff and supervisors					37

Team skills

The careful selection and thorough training of interviewers and supervisors is one of themost important steps to ensuring high quality data.

Well-trained, skilled interviewers can minimize many of the biases that are encountered in conducting quantitative surveys, particularly surveys such as these that ask people sensitive questions. Strong interviewers are able to fully understand the study purpose and questions, they can clearly communicate the information to participants and make them feel comfortable in giving honest answers, they can record the responses correctly, and they can clarify inconsistencies while the participant is still with them.
Likewise, strong team supervisors are essential to setting-up and monitoring the study process and providing quality control oversight.

The following are the important skills each of the study team members must have. Some of the skills must be screened for in selecting the team members while others can be taught during the training. Researchers at specific sites may also have other criteria for their study teams.

All **team members** must have:

- Respect for the well-being and confidentiality of the participants
- Knowledge of how HIV is transmitted and prevented
- Comprehension of the objectives of the survey and how the findings will be used
- Understanding of the consent form and questionnaire design, terminology and skip patterns
- Demonstrated ability to fully meet their responsibilities
- Prior experience conducting field surveys, whenever possible

Additionally, all **interviewers** must have:

- Ability to speak and read in the language of the participant
- Ability to put participants at ease and communicate with them clearly and comfortably

Additionally, all **team leaders and central supervisors** must have:

- Strong organizational and management skills
- Ability to select participants in adherence to the sampling frame
- Ability to trouble-shoot problems that arise in the field

In addition to the skills listed above, *central supervisors* should a strong knowledge of the geographic area of the survey and the ability to organize and move in between field teams on a daily basis. Frequently, NGO and government staff working with the populations in the survey area make good central supervisors. However, you need to account for the fact that they already have jobs and make it understood that the survey is a full-time responsibility during its duration.

If they are not identified in advance, qualified *team leaders* can be recruited from the pool of trained interviewers. If you are going to do this, you will need to select a larger number of people to be trained as interviewers to ensure you meet the interviewer quota.

Defining roles and responsibilities

Before training begins, it is necessary to define the roles and responsibilities of each of the team members. For instance, who recruits participants? Who obtains informed consent? Who secures the questionnaires and transports them at the end of the day? It is helpful to write up a list of each of the data collection steps and develop a schematic of 'data flow' with responsibilities and time frames. An example of this is shown in *Chapter 7: Survey Documentation* of this document. Without clearly outlining who is responsible for each step of the entire survey process important information - such as collecting refusal rates and recording household numbers - may be overlooked.

Training the team

An in-depth training for supervisors, team leaders and interviewers is one of the cornerstones of a good survey. While this may sound like common sense, training – just like a

careful translation of the questionnaire – is often one of the things cut short when deadlines and budgets are tight. However strong your methodologies and tools are, if the data collection team is not given enough knowledge and skills to complete its task then your survey results will be poor, and possibly unusable.

The supervisors, team leaders and interviewers should be trained as a group so they all receive the same information and understand the responsibilities of the entire team. While some aspects of the survey might be the responsibility only of the central and field supervisors and others only of the interviewers, it is important that they know as much about each other's roles as time permits. The trainer(s) should have experience in behavioral data collection, a comprehensive understanding of the sampling frame, informed consent, questionnaires, field conditions, and research ethics, and also have the ability to respond rapidly to questions that arise during training and in the field.

The amount of time for the interviewer and supervisor training may vary depending on the size and skills of the team, but usually **training takes four to five days including one to two days for field practice**. In general, the topics that should be covered include:

1. An overview of the HIV epidemic globally and locally
2. How HIV transmission occurs and what measures can be taken to prevent it and care for those already infected
3. What contributes to stigma and discrimination against people living with HIV/AIDS and their families, including exercises on what perceptions the team members might have themselves
4. Objectives of the survey
5. How the data will be used and the importance of obtaining accurate data
6. Their roles and responsibilities in the survey as well as those of the other team members
7. Sampling frame and participant selection
8. Data collection flow and schedule
9. Interviewing skills
10. Ethical survey practices including confidentiality
11. Review of the informed consent and questionnaire forms and terminology
12. Review of other fieldwork forms
13. Role playing administering consent form and questionnaire with feedback from trainers
14. Role playing of all data collection steps including household and participant selection, introduction to household, obtaining informed consent, setting up interview site, administering questionnaire, editing questionnaire, and filling out fieldwork forms
15. Field practice of all survey steps using households that are not in the sampling frame
16. Feedback on field exercise from teams and clarification by trainer(s)

Always train more interviewers and supervisors than you will need in the field. No matter how well you pre-screen the candidates, some people will prove to be more skilled at their jobs then others. Having a larger pool of people in the training means that you do not have to use interviewers or supervisors that cannot complete the skills requirements or that you

have field team members in reserve in case of illness or falling behind in the data collection timeline. However, when you train more people than you will be able to employ for the survey, make sure that your intention is clearly communicated to the trainees and that they understand that they will be on standby but not paid unless they are needed.

Important things to remember when training interviewers and supervisors:

- A quantitative questionnaire should be read without any interpretation from the interviewer. If a participant does not understand the question, that is a fault in the tool. The interviewer should not reword definitions or provide clarifications that are not documented in the questionnaire. This situation, however, frequently arises and needs to be addressed *before* field work begins with a very concise, well-translated questionnaire, fully trained interviewers, and supervisors on-site to assist.
- If you do include unprompted questions or those with 'other' as a choice, make sure you train the team carefully on how to record the answers. These types of questions may lead to interviewer bias in interpreting the answers. And they often require those responsible for data entry to try to interpret vague or illegible answers – a task that is not in their job description and often leads to even more bias.
- Informed consent needs to be done with each participant individually. It cannot be administered to a group.

Checklist 5: Team composition and training

- ☐ Determine the necessary composition for your survey team
- ☐ Calculate the number of people you will need on your team
- ☐ Define the roles and responsibilities of each team member
- ☐ Develop a list of required skills for supervisors and interviewers
- ☐ Recruit supervisors and interviewers for training
- ☐ Conduct the team training
- ☐ Select the final team members

Chapter 9

Data collection

Data collection schedule

Prior to going into the field, it is important to develop a detailed data collection schedule. This schedule will usually need to be adjusted once you begin fieldwork, but it will initially structure your target number of interviews for each field team member per day, organize your sample so that the team does not need to travel distances between households, and give you a realistic idea of whether you can achieve sample size within the survey timeframe.

There are a number of factors to consider in scheduling your data collection. The survey is attempting to interview every male and female in the household between the ages of 15 and 49 years. These are often the people that might be working, at school, or engaged in household tasks and less likely to be present or able to interrupt their work when the team arrives. It is necessary for the supervisors to determine before data collection begins when is the optimal time for reaching participants, and to keep in mind that these times may be different for the populations residing in the camps and those in the communities. They should also build into the survey schedule adequate time for revisiting households in case of absence or requests to interview at a different time.

Another issue you might encounter is mobility of household members, whether for short or long periods of time. If a person is away from the household but will be returning during the period of the survey, then an appointment should be made with the head of household to return to the home when the participant will be available. If the participant will be gone during the entire data collection period, then they will be coded as absent. In populations where there is a great deal of long term travel, then you may want to consider this in defining the recruitment criteria and alter the population definition to encompass only those people who have been residing in the household in the past two weeks or past month.

Household selection and recruitement

Based on your sampling frame and data collection schedule, a fixed number of households should be selected each day by the supervisor. When the survey team reaches the selected household, the supervisor should introduce him/herself and the study team, clearly explain the objectives of the survey, what the information will be used for, that the survey is anonymous and participants' confidentiality will be upheld, how long the questionnaire will take, and obtain permission from the head of the household for the interviews. This is not the time to administer the informed consent form. First, the household must choose to participate, then each individual should be given the opportunity to agree or decline him/herself.

Household refusals and absences and abandoned households

If the household is unable to participate at that time but would be willing to at another occasion, the supervisor should schedule a suitable time for the team to return. If a household is absent, the supervisor should record this information and determine another time to return. The team should return to an absent householdup to three times if it is logistically feasible within the time, distance and budget constraints of the survey. If the team cannot recruit the household after multiple attempts, this information should also be recorded.

If a head of household refuses to participate, then it will be considered a refusal and this information should also be recorded. Whenever possible, collect information on the household to assess whether it has different characteristics from the households that elected to participate. **No replacement household should be selected for refusals or absences** because this may introduce sampling bias and an appropriate non-response rate has already been built into the recommended sample size to account for these eventualities.

A household will be considered abandoned if neighbors report that nobody has lived in that household for more than one month or if the inhabitants have repatriated. This household should not be replaced by another household and it should be reported as abandoned. During periods of repatriation, there may be a high number of abandoned households in the refugee villages. It is crucial that you know about the possibility before data collection begins, and where necessary inflate your sample size to account for a large number of absent households *(for more information, see Chapter 5: Sampling)*.

Participant selection

Once the household has agreed to participate, the team leader will work with the head of household to develop a list of all household members by name, age, sex and relationship to the head of household. The team leader will then identify all of the men and women in the household who are eligible for recruitment.

Participant recruitment

The interviewers should set up a private, quiet place to administer the questionnaires where they will not be overheard or interrupted. Each person recruited should be read the informed consent individually in this private place and given the opportunity to either accept or refuse involvement in the survey. More information on obtaining informed consent is covered below in *Chapter 6: Questionnaire and informed consent* and *Chapter 14: Research ethics.*

Participant refusals and absences

If the individual does not want to participate in the survey, this information should be recorded. For people who are unsure, the interviewer can answer additional questions about

the survey and reassure them about the confidentiality of their answers, but they should not attempt to coerce them into participating, nor should they allow any other household member to do the same. Individuals who refuse to participate should not be replaced, and the refusals should be recorded.

If a household member is absent at the time of the survey, the supervisor should record the details of this participant and discuss with the household when the best time and place would be to schedule the interview. The team should attempt to recruit the participant up to three times. If they are unsuccessful after multiple attempts, the individual should be recorded as an absence. Absent participants should not be replaced.

Understanding why households and individuals refuse to participate

Before building an estimated non-response rate into the sample size and training the teams to recruit participants, the investigators should understand the factors in the communities that might influence response rates. For instance, one complaint of population members may be the number of surveys they are asked to participate in without ever seeing any results from their involvement. The investigators must be prepared to disseminate and use the data to benefit the populations, or they should not be doing the survey in the first place. How the information will be used needs to be clearly communicated during recruitment, and then the team should follow through with its data use plans.

Recording information on why people refuse to participate can be valuable to understanding if the results of the surveys will be biased, better protecting the rights of the participants, and developing stronger recruitment measures for further surveys (such as greater involvement of community members during the survey preparation stage).

Recording household and participant information

Often information about household and participant recruitment, absences, refusals and household abandonment is not collected by the field teams. There are a number of reasons for this including the team not having the proper forms or not understanding how to fill them out, or the forms being considered just another burdensome piece of paperwork with no added value to the survey. However, the **data collected on refusals and absences is important**. The information allows the supervisors to better manage the sampling and data collection process, to monitor refusals and absences, and to respond to issues encountered during fieldwork - such as very high refusal rates - while there is still time to make changes. The information also allows the people analyzing the data to assess whether the sampling methodology was followed correctly, whether the sample size was adequate to analyze for

key indicators, and what potential biases might have been introduced by large numbers of absences or refusals.

During training, the forms should be clearly explained and the teams should understand how the information is going to be used. There should also be one person on the team responsible for managing the information, preferably the team leader. Figure 6 below is an example of the household information sheet. The form does not need to be set up in this exact way, but the information on it does need to be collected. In order for the supervisors and team leaders to take greater ownership in managing the sampling data, it may be useful to allow them to develop a layout of their own which they find understandable and easy to use. An example **household information sheet** is provided for you in **Appendix 1** of this guide.

Figure 6: Example of household information sheet

Household information sheet (daily summary)

To be completed by team leader and returned to central supervisor

Serial number of household	Number eligible people (15-49) in household	Number participants recruited	Number participants refused/ absent	Reason for household and participant refusal/absence	Household absent			
					Date visit 1	Date return visit 2	Date return visit 3	Absent household recruited
201	4	4	0	N/a	Nov 3			
202	3	0	3	Household head did not want to 'waste time'	Nov 3			
203	Unknown	0		Household repatriated	Nov 3			No
206	5	5	0		Nov 3	Nov 4		Yes
215	0	0	--		Nov 3			
216	4	4	0		Nov 4	Nov 4	Nov 5	Yes
217	4	3	1	Participant was ill	Nov 4			
219	3	3	0		Nov 4			

Figure 7 below provides an example of the information collected in the participant information sheet including household and participant identification number, age, gender, relationship to the head of household, eligibility to participate, refusal, absence, and whether the interview was fully or partially completed. Just as with the household information sheet, the teams need to be trained on how this form should be filled out and how the information in it will be used. There should also be a single person responsible for managing the

information on the sheet, preferably the team leader. A full **participant information sheet** is provided for you in **Appendix 2** of this guide.

Figure 7: Example of participant information sheet

Participant Information Sheet

One household per sheet. Do not put multiple households on the same sheet.

To be completed by team leader for every household and submitted with completed questionnaires

Serial number of household	Household member number	Age (yrs)	**Gender** 1. Male 2. Female	**Relationship to the head of household** 1. Household Head 2. Spouse 3. Son/ Daughter 4. Father/ Mother 5. Brother/ Sister 6. Other relative 7. Living in household but not a relative	Visit 1 Visit 2 Visit 3 1 = Refusal 2 = Individual not eligible 3 = Questionnaire completed 4 = Questionnaire partly completed 5 = Household member absent 6 = Others (Specify) **(for each household member record the correct answer)**
217	1	42	1	1	3
	2	37	2	2	3
	3	30	1	5	1
	4	18	2	3	3
	5	12	2	3	2
	6	6	2	3	2
	7	1	1	3	2

Setting up the interviewing space

One important aspect of recruiting participants and getting candid responses to the survey questions is the assurance of confidentiality regarding their identity and their answers. It is not enough that personal identifiers such as names are not recorded. The participant must also feel secure that their answers during the interviewing process are not overheard or monitored. A private space should be found where no one will be able to listen in and there will be no distractions during the interview, and no one should be allowed to enter the space during the length of the interview. Making sure you have an appropriate interviewing space will help to put participants more at ease and increase the likelihood that they will feel comfortable in answering the questions.

Obtaining informed consent

Prior to administering the questionnaire, informed consent needs to be obtained from each individual. This should be done by the interviewer or other assigned team member. Informed consent should take place in the same private space as the interview and not in a public area or with a group of people. The minimum amount of information needed for informed consent has been provided in the questionnaires in Appendices 3 and 4, and the person responsible for asking for consent can clarify additional questions the participant might have. However, the interviewer should not answer questions regarding HIV transmission at this time because this will influence the participant's responses to the questionnaire. After the questionnaire is completed, the interviewer or supervisor should answer all of the participant's questions regarding HIV in full.

More detail on the informed consent process is provided in *Chapter 6: Questionnaire and informed consent.*

Administering the questionnaire

After the survey space has been set up and informed consent has been obtained and documented, the interviewer can administer the questionnaire. Details on this process are provided in *Chapter 6: Questionnaire and informed consent.*

Although a participant may have given consent before the interview began, this does not mean that they are obliged to complete the interview. Sometimes a participant may decide to stop part way through the interview. This may happen because they do not feel the space is private enough to answer the questions, they are uncomfortable with the nature of questions, or they need to return to their work. If they do want to end prematurely, the interviewer should establish the reason and attempt to rectify it or reschedule the interview. However, it is the right of the participant to stop the interview at any time, and they should not be forced to complete it. Many of the issues that may cause people not to complete interviews can be dealt with prior to administering the questionnaire by choosing an appropriate interview site, clearly explaining the objectives of the survey and amount of time it will take, and recruiting households at times that are convenient for the participants (or, failing that, scheduling interviews at better times).

Another issue that may be encountered is that while an individual agrees to participate, he or she does not answer the questions fully or (maybe the interviewer feels) honestly. It is not the interviewer's role to question the veracity of the participant during the interview process. What the interviewer should do is look at the questions that are being answered for internal consistency and clarify the answers with the participant. For instance, if a participant responds that she has been sexually active and is currently married, but in the later section of the questionnaire she says that she has had no sex partners in the past 12 months, the interviewer should clarify the potential inconsistency between the different answers. What the interviewer cannot do is claim that the participant is lying or record any information other than what the participant provides.

The sample size of the surveys is designed to be robust enough that individuals who do not want to provide answers to some questions should not skew the analysis of the population as a whole. Remember that an objective of these surveys is to understand behavioral risks in a *population*, not in a single person.

A larger problem enters when there appears to be systematic non-response to certain questions. For instance, a majority of unmarried youth may respond that they have never had sex when other surveys and field research may have previously identified that the opposite is true. Behavioral surveys similar to the ones outlined in this manual have been done on all continents and many different populations. While often the questions are sensitive and some people will chose not to answer them or minimize their risk in their responses, on a whole most people who have agreed to participate will also attempt to answer the questions as fully as possible. Factors that may influence systematic non-response need to be clearly understood at the design stage of the survey, and steps should be in place whenever possible to minimize them. If these issues seem insurmountable, the investigators need to question whether the objectives of the survey can be fulfilled and whether conducting the survey will provide them with unbiased data that can be used to benefit the populations.

When a survey that is already in the field begins experiencing systematic non-response, the investigators must internally discuss what they are finding and explore with community members what might be influencing answers. Perhaps young people are not comfortable answering questions in a site where their family is present, even if the interviewing space is private. Perhaps it is taboo in the community to have had sex outside of marriage, and no matter what lengths the team takes, this taboo cannot be overcome. These are vital circumstances to understand the survey begins. Any changes made in survey methodology after field work begins may introduce bias into the survey results. The teams must weigh whether to carry on with their established methods and have higher than expected non-response rates or to make an environmental change, such as interviewing all youth outside of their households, and the possible biases that could be introduced with these changes. In all cases, this information *must* be documented so that it can be considered at the analysis and interpretation stages of the survey.

However, do not automatically assume that, for instance, because a young person reports that they are not having sex, they must be lying. Look at other available data and talk to community members and the young people themselves. Often there are anecdotal perceptions of sexual behavioral among youth (and other populations) that are not representative of the youth population as a whole. One of the benefits of quantitative surveys is to enumerate a group's behavioral risk factors and to steer away from the common practice of generalizing the experience of a few people to everybody in the population.

❧ Reviewing the questionnaire

Once the interview has been completed, and while the participant is still present, the interviewer should review the questionnaire for completeness and consistency of answers. The interviewers also need to make sure that they recorded the answers clearly and that the nu-

meric code for each answer is written in the box to the right of the question. This review process may take a bit of time when the interviewers are still getting used to the questionnaire, but they will soon be more comfortable with the tool and able to complete it more rapidly.

Editing the questionnaire

After the interview is completed, the questionnaire should be given to the team leader who will review, or *edit*, it. The team leader checks to make sure that the questionnaire has been filled out completely, that there are no mistakes in skip patterns, that the answers are consistent, and that the interviewer has recorded the answers legibly in the boxes provided. If the questionnaire is not correct, the interviewer should return to the participant to complete it while he or she is still at the household.

Why record the responses as a number?

The reason that the responses to the questions are recorded as a number and not just circled on the questionnaire is to assist the data entry clerks in their job. The people entering the data should be able to run their eyes directly down a line on the page and enter the numbers that correspond with each question. It slows them down if they need to determine what answer has been circled for each question, and it also introduces the potential for transcription error.

Consider the examples below. In the first, the interviewer has circled the answers to the questions. In the second, the interviewer has written the number of the answer into the spaces provided. If you were entering the data, which method would you prefer? Which one is more open to transcription errors? Look at question 104. As a data entry clerk, is it clear to you which choice has been circled?

Figure 8: Circling the responses on the questionnaire

101.	Have you ever been married?	(1) = Yes 2 = No	|___|	If **NO** go to **120**
102.	How old were you when you first married?	Age in years 17 99 = Don't Know	|___|___|	
103.	What is your current relationship status?	(1) = Currently married 2 = Never married 3 = Divorced 4 = Widow/ Widower	|___|	If **NOT CURRENTLY MARRIED** go to **122**

104.	Are you in a monogamous or polygamous marriage?	1 = Monogamous 2 = Polygamous	|___|	
105.	Are you currently living with a long-term partner?	1 = Yes 2 = No	|___|	

Figure 9: Coding the responses as a number

106.	Have you ever been married?	1 = Yes 2 = No	|_1_|	If NO go to 120
107.	How old were you when you first married?	Age in years 17 99 = Don't Know	|_1_|_7_|	
108.	What is your current relationship status?	1 = Currently married 2 = Never married 3 = Divorced 4 = Widow/ Widower	|_1_|	If NOT CURRENTLY MARRIED go to 122
109.	Are you in a monogamous or polygamous marriage?	1 = Monogamous 2 = Polygamous	|_2_|	
110.	Are you currently living with a long-term partner?	1 = Yes 2 = No	|_2_|	

During editing, the team leader should also check all of the write-in responses in the questionnaire, such as spaces provided to indicate an answer other than those prompted. The team leader should make sure that the 'other' cannot be recoded into one of the pre-existing answers. In the example below, the participant said the last time he visited the host community was to go to church, and the interviewer has written this response as:

11= Other (specify)__Church___

However, this question already has the option 10= 'attend religious service'. The team leader should reclassify the answer to option 10.

111.	The last time you visited the refugees/ host community, what was your reason? **Only one answer can be recorded**	1 = Employment 2 = Trade 3 = Shopping/ Market 4 = Health care 5 = School 6 = Entertainment |___|___| 7 = Food 8 = Visit relative/friend 9 = Collect firewood 10 = Attend religious service 11 = Other (specify) _**Church** _____ **Note: Responses should be revised to include other locally relevant reasons**

Any problems that arise should be corrected while the team is still at the household, not in later on when the data is being entered and the ability to make the changes is lost. This process will also give the supervisor a chance to monitor the understanding and skills of each interviewer and provide additional training to him or her when problems are encountered. Central supervisors who are monitoring the survey teams should also do random checks of the questionnaires for accuracy on a daily basis.

Thanking participants

After completing the interview, each person and household should be thanked for providing their time and assistance. The team leader should answer any additional questions that the household might have including those about the survey itself and questions regarding HIV and STI prevention, care, and services. Investigators should also give their contact information in case participants have additional questions to ask at a later date.

Collecting and securing field work forms

At the end of each day one person, the central supervisor should be responsible for collecting all of the questionnaires and fieldwork forms and transporting them to a central area. The questionnaires should be locked in a safe place and not kept in a person's home or on his or her desk. Doing this each day will ensure that the questionnaires are not accessible to anyone outside of the field team while the identity of the households and participants can still be traced. It is also a precaution to questionnaires being misplaced.

Team debriefing

Frequent team debriefings, particularly in the early stages of data collection, are an important step to keeping the survey on track and trouble shooting problems early. During the first week, these debriefings can be held daily. The debriefing is a forum for all team members to discuss what problems they are encountering and how they are handling them, whether they are able to keep to the pre-determined schedule, what are the plans for the following day, and other topics. As data collection continues, the supervisors might chose to hold less frequent debriefings depending on how the teams are progressing.

Checklist 6: Data collection

- ☐ Develop a data collection schedule
- ☐ Recruit household based on the sampling frame
- ☐ Determine who in household is eligible for participation
- ☐ Recruit individuals
- ☐ Record household and participants information including refusals and absences
- ☐ Schedule a time to return for absences
- ☐ Set up the interviewing space
- ☐ Obtain informed consent
- ☐ Administer the questionnaire
- ☐ Review the questionnaire
- ☐ Edit the questionnaire
- ☐ Thank the participants
- ☐ Collect and secure fieldwork forms
- ☐ Debrief the team

Chapter 10

Data management

During the design and implementation of a BSS, data management may be the least considered step in the entire process. Frequently, survey staff are instructed to 'enter and clean the data', with no additional guidance on how this should be done. However, there are a number of biases that can be introduced into the survey at this stage (discussed in the sections below) which can be minimized by trained staff and carefully laid out data management processes. It is important not to underestimate the amount of time needed and staff skills necessary for data management. Often it can take longer than the fieldwork itself, but it becomes rushed as deadlines near. When this happens analysts later find serious problems within data sets, and rectifying these issues is a long, and sometimes unfeasible, process.

What is data management?

Data management actually begins well before data cleaning. It defines how the survey information is collected and moved from each stage to the next. In the previous chapters we discussed how the interviewer administers and reviews the questionnaire, and then passes the questionnaire to the supervisor for editing. These are all steps in data management; first reviewing the data and correcting errors and then organizing it in a way that is easiest to use when it is turned over to the next person responsible in the survey chain, in this case the data entry clerk. In this manual data management in the field is discussed separately from the data management needed from data entry onward because different people are usually responsible for the different areas.

So why not just call it 'entering and cleaning the data'? As discussed later in this chapter, data management is more than just cleaning the data. It often requires combining data sets, re-labeling variables, or other steps that streamline the analysis. In essence, data management is all the things you need to do to prepare the data for analysis.

Choosing the data entry software

There are many different types of software available for entering data from simple spreadsheet programs to more sophisticated packages. In the past, EpiInfo (both DOS and Windows versions) and CS Pro have used to enter survey data because they are free and more

easily accessible in the field. CSPro is often used by national and regional statistics bureaus, and if you are recruiting data managers and entry clerks from these offices they may already be familiar with the software. It can be downloaded at http://www.census.gov/ipc/www/cspro/index.html. In addition, a program called EpiData Entry has been used in the field with positive results. EpiData Entry is freeware, and it is recommended for data entry and management. The program and instructions manual can be downloaded at http://www.epidata.dk/.

You do not have to use the same software for data entry and management that you will use for your analysis. Most database software will allow you to export your data in different formats which can then be imported into your analysis software. There are also specialized statistical transfer packages such as DBMS Copy™ and StatTansfer™ that make moving datasets between software programs relatively painless.

Designing a data entry program

The data entry form should be designed to match the questionnaire . Do not, for example, change the order of the questions or separate single response questions into multiple variable fields. If a response to a question is answered in months and years, the data entry program should have separate fields for months and years. For questions that have '99' as the response for 'Don't know', the field should allow two numbers to be entered.

CSPro and EpiData Entry have a step-by-step tutorials explaining how to design a data entry program including building in consistency checks and skip patterns. Designing a data entry program can be quite straightforward, but there are several areas that can be overlooked or cause confusion and are discussed in more detail below.

In Appendix 5, there is a **Coding and Analysis Guide** for the expanded questionnaire that is a helpful reference for designing your data entry program. However, make sure that any differences between the coding guide and your questionnaire are reflected in the database.

Consistency checks and skip patterns

Most data entry software allows the programmer to build in consistency checks to make sure that the entered data is not outside the bounds of what is allowable in the survey. For instance, if the inclusion criteria for your survey is women and men between the ages of 15 and 49, you can set ranges for the variable 'age' that will not let a numeral lower than 15 and or higher than 49 to be entered. If the data entry clerk types in '14' for an age, the program will not allow him or her to continue until it has been corrected. Many of the questions in your survey will only have '1' (yes), '2' (no), '98' (no answer) and '99' (don't know) as answers. The data entry program should be designed so that no other number can be entered.

Some software programs also have a feature that will let you build in consistency checks between responses. For instance, if a respondent reports that he is married and living with his spouse, but then when asked if he has had any regular sex partners in the past 12 months he

reports 'no', the program will flag the answer so that the data entry clerk knows to go back to check that all of the responses were entered correctly. Similarly, if a participant answers that she has never had sex, but later replies that she had a casual sex partner in the past 12 months, the program will alert the enterer to the discrepancy.

Another important tool in doing data entry is building the skip patterns directly into the data entry program. Then if a respondent replies that he has never had a transactional sex partner, the program will skip over the rest of the questions in the section and move the program to the next appropriate question, lessening the time the data entry staff needs to scroll through the sections as well as minimizing the data entry error from not following the correct skip patterns.

As discussed previously, the questionnaires provided in the appendix have very few multiple response questions, meaning a single question where a respondent can give more than one answer. The reason for avoiding multiple response questions is that they are often difficult for the interviewer to fill out correctly, they increase the length of the survey, they make the coding and data entry process more time consuming and the data sets much larger, they are difficult to analyze, and frequently the data is not used. Investigators generally want to keep their surveys shorter so that they take less time to administer, enter and analyze. But surveys with many variables are also problematic if you are using EpiInfo (both DOS and Windows versions) for data entry, because the program restricts the number of variables in the data entry program to less than 255. EpiData Entry and CSPro does not have any restrictions in the size of the dataset.

One exception where multiple response questions are used in the questionnaires in this manual is for those measuring exposure to interventions. The reason multiple response questions have been used here is because a great deal can be learned from understanding every source of information and services available to the population, and the participant should not be restricted to naming only one of their sources of HIV information.

Figure 10: Example of a multiple response question in the questionnaire

602.	From what sources have you received information on HIV/AIDS in the past 12 months? **Unprompted question. Record all answers given**	*Mass media* 1 = Radio |___| 2 = TV/ Video |___| 3 = Newspaper |___| 4 = Poster/pamphlet |___| *Health services* 5 = Health facility |___| 6 = VCT center |___| 7 = ANC/PMTCT center |___| *People* 8 = Community health worker |___| 9 = Friend |___| 10 = Family member |___| 11 = Person living with HIV/AIDS |___| 12 = Peer outreach worker |___| *Other places* 13 = School |___| 14 = Place of worship |___| 15 = Public meeting |___| 16 = Others (specify) _____ |___| *Note: Other locally relevant responses should be added including specific interventions being conducted in the camp/community.*	

In order to enter the above example correctly, each possible answer should be given its own variable field. '*From what sources have you received information on HIV/AIDS in the past 12 months*' will become 15 different variables in the data set, and each participant will have a 'yes' or 'no' answer for every one of the variables.

Denominators

Skip patterns in the questionnaire mean that not all questions will be answered by every participant. The data entry clerks should enter only the information provided to them in the questionnaire. They should not fill in 'no' for people who have been skipped in the questioning or record in the database any information other than what is written on the questionnaire and other relevant field forms. If no information is recorded for a question, the data entry program will identify it as 'missing'. Later in the data management and cleaning process, some of those 'missing' answers may be recoded back into the denominator, but that should not be done during data entry. How to determine the correct denominator for analysis will be discussed later in this chapter and in *Chapter 11: Data analysis.*

Open questions and write-in respons

The behavioral surveys described in this document are quantitative, meaning every response needs to be quantified, or counted up, so it can be analyzed. A majority of the time, each possible response to a question will be given a pre-determined numeric answer to be entered into the database. However, sometimes questionnaires will have 'open' questions or 'other' responses where the answer is written in by the interviewer. As discussed in *Chapter 6: Questionnaire and informed consent*, write-in answers can prove very difficult to manage and analyze. When they cannot be avoided, they must be carefully controlled.

The lead investigator should try to develop a list of numeric codes for each of the possible write-in answers, and these codes should be assigned to the answers before data entry, usually when the supervisors are editing the questionnaire.

Sometimes it is difficult to make a list of numeric codes for each possible response. In these instances, the response may be written-in by the interviewer and then entered into the dataset in text (also known as a 'string variable'). However, using string variables can introduce errors into your data set, and every effort should first be made to assign a numeric value to each response.

Write-in responses are frequently used is when an option is given for 'other' in a single response question. Frequently, the written-in 'other' response actually fits into one of the preassigned answers. It is the responsibility of the team leader to review these responses and recode them when necessary. See *Chapter 9: Data collection: Editing the questionnaire* for more information on coding 'other' responses.

When an 'other' answer is different from the all the numeric answers given in the questionnaire, there are several alternatives for how to manage the information. Most often, data entry clerks will enter the numeric code for 'other' and the written-in information will not be

used. When an 'other' response is infrequent, this may not influence the interpretation of the data. But when many people are giving the same 'other' response, you lose the ability to analyze the information by not entering it into the dataset. After all, why ask the question if you are not going to use the information?

A second option is to enter the numeric code for 'other' in the original variable field and then create a separate variable in the database to enter the various responses. However, in doing this you are creating a text (string) variable that may be difficult to analyze. Take the example in Table 4 below. The first table shows the responses to the question *'What is your religion?'*,The second table illustrates the variety of ways the data was entered into a second variable for the 'other religion' responses to the question.

Table 4: Example of a write-in response for 'other' religion

What is your religion?

	Freq.	Percent	Cum.
Catholic	1,471	43.92	43.92
Protestant	1,180	35.23	79.16
Moslem	509	15.20	94.36
Other (Specify)	189	5.64	100.00

Written-in responses to 'Other religion'

.....	Freq.	Percent	Cum.
ATHEIC	6	3.21	3.21
ATHETIC	8	4.28	7.49
LEGION OF MARY	5	2.67	10.16
LEGIONS MARY	1	0.53	10.70
LEGIONS OF MARY	2	1.07	11.76
LEGO MARIA	1	0.53	12.30
MAM EKANISA	1	0.53	12.83
MAN EKANISA	1	0.53	13.37
NO RELIGION	1	0.53	13.90
NON BELIEVER	6	3.21	17.11
NON RELIGIOUS	5	2.67	19.79
NONBELIEVER	1	0.53	20.32
NONE	25	13.37	33.69
ORTHODOX	22	11.76	45.45
PAGAN	64	34.22	79.68
PCEA	1	0.53	80.21
SABATH	1	0.53	80.75
atheic	2	1.07	81.82
does,nt have	1	0.53	82.35
doesn,t have	1	0.53	82.89

legion	1	0.53	83.42
legion mary	10	5.35	88.77
legion of mary	1	0.53	89.30
nine	1	0.53	89.84
none	6	3.21	93.05
orthodox	4	2.14	95.19
pagan	4	2.14	97.33
pegam	1	0.53	97.86
religion of mars	2	1.07	98.93
relion of mars	1	0.53	99.47
sabath	1	0.53	100.00
Total	**187**		**100.00**

Far less time consuming then trying to decipher and analyze the information in the above table is to find out about the different religions practiced the area before beginning the survey and to include those religions in the original list of choices. If you do choose to write-in the answers in a separate variable, it helps to have a central data manager to keep a running list of the different responses and instructions on the exact way they should be typed including the spelling, capitalization, and abbreviations that should be used. This whole process is much easier when the people doing data entry are in a central location, preferably in the same room, and working under close supervision.

It is also possible to create numeric codes for the 'other' responses as the questionnaires are being edited or when they are encountered by the data entry clerks. This process entails creating a running master list of 'other' responses for each question that is managed by a single person and accessible to all the data clerks. The 'other' responses can then be entered as a second variable and merged with the original *'What is your religion?'* variable during data management. However, this system can be quite difficult to manage with a large number of questionnaires and multiple data entry clerks, especially when they are working in different places or if there isn't a single person responsible for overseeing the coding.

The decision about how to manage written-in responses falls to the lead investigator of the survey and will depend on how the survey and team are structured and supervised. Ultimately though, good survey preparation and thorough field editing should minimize the number of written-in responses that will need to be managed.

Information to include in the data entry program

As explained previously, the data entry program should follow the structure of the questionnaire exactly, and it should include all of the questions in the questionnaire. The following information from the field work forms should also be entered into the database:

1. **All information in the identification section**

Questionnaire serial number |___|___|___|___|

Household serial number |___|___|___|___|

IDENTIFICATION	
COUNTRY	|___|
REGION/ PROVINCE	|___|___|
CAMP/ SURROUNDING AREA (Camp = 1, Surrounding area = 2)	|___|
IDENTIFICATION NUMBER OF CAMP/ VILLAGE	|___|___|
CLUSTER NUMBER	|___|___|
URBAN/ RURAL (Urban = 1, Rural = 2)	|___|

2. **Code of interviewer**

NAME AND CODE OF INTERVIEWER

_____ |___|___|___|

3. **The code for the controller at each stage and data entry** clerks should be entered, but it is not necessary to enter the date.

CONTROL				
	CONTROL ON FIELD LEVEL	CONTROL IN CENTRAL OFFICE	DATA ENTRY CLERK 1	DATA ENTRY CLERK 2
NAME				
DATE	----------|__|	----------|__|	----------|__|	----------|__|
	----------	----------	----------	----------
REMARKS				

4. Date of interview and start and end time of interview

Date of interview: __/__/ day /__/__/month

Start of interview: __/__/ h /__/__/min

End of the interview: __/__/ h /__/__/ min

5. Household information

HEAD OF HOUSEHOLD
PARTICIPANT IS HEAD OF HOUSEHOLD OR REPRESENTATIVE OF HEAD OF HOUSEHOLD ..... |___| 1 = Yes 2 = No 3 = No head of household or representative present

NUMBER OF PEOPLE IN HOUSEHOLD
Total number of people living in household .....|___|___| Total number of eligible people aged 15-49 living in household .....|___|___|

6. Result of household recruitment

RESULT OF HOUSEHOLD RECRUITMENT
RESULT OF HOUSEHOLD RECRUITMENT ..... |___| 1 = Head of household agreed to household participation 2 = Head of household refused household participation *Reason for refusal* ..... 3 = Household not eligible 4 = Household temporarily absent Date and time of first visit ..... Date and time of second visit ..... Date and time of third visit ..... Reason for household's absence ..... 5 = Household abandoned 6 = Household on extended travel 7 = Other (specify) .....

7. Participant's relationship to head of household and result of participant recruitment

RESULT OF PARTICIPANT RECRUITMENT

RESULT OF PARTICIPANT RECRUITMENT .....|___|
1 = Agreed to participate and interview was completed
2 =Agreed to participate but interview was not completed
3 = Refused to participate
4 = Absent
Date and time of first visit .....
Date and time of second visit .....
Date and time of third visit .....
Reason for participant's absence .....
5 = Other (specify) .....

RELATIONSHIP TO HEAD OF HOUSEHOLD

PARTICIPANT'S RELATIONSHIP TO HEAD OF HOUSEHOLD .....|___|
1 = Head of household
2 = Spouse
3 = Son/Daughter
4 = Father/Mother
5 = Brother/Sister
6 = Other relative
7 = Living in household but not a relative

Training data entry clerks

The data entry clerks are vital members of your survey team. If they did not participate in the original training and fieldwork, then it is important that they become familiar first with some of the general aspects of the survey including its objectives and how the data will be used. They should be trained in the structure of the questionnaire, and in particular the skip patterns and coding. They should then learn about the structure of the database and how to enter the data into it, with a particular emphasis on the process for entering write-in responses.

Developing a data entry schedule

Just as with the fieldwork, it is helpful to develop a data entry schedule so that the clerks know how many questionnaires they are expected to enter each day and how long the proc-

ess will take. The length of your questionnaire, the accuracy of the editing, and the skills of your clerks will influence how many questionnaires they will be able to enter each day. The schedule may be adjusted as the data enterers become more experienced with the job or as unplanned events crop up, however it is often useful to develop a (flexible) structure at the outset so that people know what is expected of them and in what timeframe.

An experienced data entry clerk should be able to input at least 50 BSS questionnaires a day. However, it will take some time for the clerks, regardless of experience, to become comfortable with the questionnaire and data set format. You should expect that they will take a number of days to get up to full speed, and many will never reach that number.

Data entry clerks are often paid a fixed salary per day in the same manner as the rest of the survey team. However, paying clerks per questionnaire entered offers them an incentive to work more quickly. If the latter method of payment is used, the data entry should be monitored closely to ensure that accuracy is not being lost for the sake of speed, and a questionnaire should not be considered completed until it has been entered completely and correctly.

Data entry and supervision

To ensure the highest accuracy in your data entry, it is always recommended that you do **double data entry**. This means that each questionnaire is entered into a separate dataset by two different people independently. Most data entry programs, including EpiInfo, CSPro and EpiData Entry allow you to append the two datasets together when they are completed in order to identify the entries that have discrepancies so that they can be easily corrected.

Double data entry may strike some people as a waste of effort. Why do something twice when you only need to do it once? But double data entry can increase the quality of your data set and allow those responsible for data cleaning to find the inaccuracies in the data entry immediately. If you find you have a poorly entered data set or a data entry clerk that didn't fully understand their job *after* data entry has been finalized then it often becomes the task of the analyst to go back through the questionnaires to make the needed corrections. If there is systematic data entry error it may be necessary to re-enter some, if not all, of the problematic questionnaires. This process can take considerable time and is often left to supervisors and analysts who have many other competing priorities. Close supervision of data entry is always good practice, and it becomes even more important if you are doing only a single entry of each questionnaire.

It is not the responsibility of a data entry clerk to guess what an unclear code might be or to make decisions how to enter incorrectly filled out questionnaires. There should always be a supervisor who was in the field available to answer these and other questions. If a supervisor is not immediately available when there is a question about a write-in or other response, the number in question should be marked and noted, and the supervisor should help the clerk to complete the entry as soon as he or she is available.

Supervision means not only training the data entry clerks, monitoring their progress and responding to problems as they arise, but it also entails frequently cross-checking the data entry fields against the questionnaires to make sure they are being filled in correctly. Ideally, you should do as many checks as possible to find errors and identify data entry clerks who may require additional training and supervision. Usually, supervisors will try to check about 10% of the questionnaires, but particularly when you are not doing double data entry it is better to check even more questionnaires while they are still being entered and can be more easily corrected. Having to comb through thousands of questionnaires when it comes time to clean the data is not a pleasant task, and often valuable data will end up being discarded because incorrect entries cannot be fixed. This can also introduce bias into the survey, particularly if there were systematic errors in the data entry.

In order to manage the data entry process including overseeing the data entry clerks, track and organize the questionnaires, running consistency checks, and combining and cleaning the data, it is highly recommended that a full-time data entry manager be hired for these surveys. It is difficult for the lead investigator or central supervisors to manage this process on top of their other duties. Data entry and management that is done poorly will result in low quality data and lengthy delays in disseminating your data.

Combining the data sets

After all of the data entry is finished, the data sets should be appended together. This means that all participants, both refugee and host community, should be in a single data set. It is much easier to do analysis of only one data set, particularly when you are looking for differences between populations. However, you may want to consider keeping datasets from different geographical locations separate.

Cleaning the data

Data cleaning should be a *systematic* process. Often people immediately jump directly from data entry into their analysis and correct errors as they come across them. This non-systematic process means that you will need to correct your data set each time you find an error (which may necessitate finding the questionnaires), revise the analysis that you have already done, and worry throughout the analysis whether problems are lurking in the data set that you have yet to (or may never) discover.

Data cleaning also does not entail simply checking a few random variables to make sure they look okay and starting to analyze. As with all the other steps in these surveys, you need to have a plan.

Step 1:
If you have done double data entry (as highly advised), first append both of the data sets together. Most data entry programs will allow you to screen for differences between the data entry fields in both copies. Identify each of the variable responses that are different in the

two data sets for the same participant. Sometimes it will be clear which data set is incorrect. For instance, '5' will be entered as the participant's age in one data set, and '25' will be entered in the second. Other times it will not be as clear, and you may need to reference the questionnaire to identify the correct response. Choose one of the data sets to be the 'master' data set (preferably the one with the least mistakes), save it as a new file and make the changes into the new data set. Do not make changes directly into either of the data sets you are comparing or you will lose your original files and will not be able to reference them later if necessary. A note: This step in the cleaning process is only correcting for data entry error, it is not cleaning out any mistakes made by the interviewers or team leaders that would be common to both data sets.

Step 2:
Review your data set for completeness. You want to determine whether all of the questionnaires were entered, that they were entered only one time, and that each participant has a unique identification numbers in the data set. Tabulate the questionnaire identification number and make sure that the total is equal to your sample size, that each number was entered only once, and that the questionnaire identification numbers listed are the same as those in your survey. The same process should be done for the household identification numbers to ensure that all households selected are in the survey, the number of participants for each household is correct, and more than one household was not assigned the same number.

You will also want to tabulate other characteristics such as gender, age and whether they are a member of the refugee or host member to make sure the number of people in your data set is the same as in your sample.

Step 3:
Screen for outlying variables. Outliers are numbers in your data set that are not within the possible or plausible range of answers, or they may be correct answers that could skew your results. Sometimes outliers will be obvious. If possible responses to a question are 'yes' (1), 'no' (2), 'don't know' (99), then the number '7' has no place in that variable. If the age range of your population is 15-49, then there should not be an 83-year-old in your data set. A data entry program that specifies what numerals, or numeral ranges, can be entered into a variable field will have resolved many of these outlier problems before the data cleaning stage, as should double data entry. To correct the outliers, you can either reference the questionnaire or set the outlying values to missing.

Sometimes, however, an outlying answer is less obvious, and it may indeed be the correct one. Take the question, *'What was the age of your most recent casual partner?'* A young women of 18 might answer '65-years-old', another woman of 19 might say '62-years-old', and two other young women may report partners in their late 50's. These may in fact be the ages of their last casual partners, but when analyzing the mean age of the last casual partner for women aged 15-19, they could skew the age up because they are far older than the median age of, say, '29'. In this situation, the original data should be retained, however the outliers should be noted and both median and mean ages should be calculated for that variable to discern how the outliers are affecting the interpretation of the response.

Step 4:
Check your denominators. Many of the questions in the survey should be answered by the total population. Look at each of those questions and make sure that the total number of people who answered the question is equal to the total sample size.

While some questions in the survey are answered by everybody, others may only be answered by a specific sub-group of participants. In this case, the denominator of your variable should not be the same as the total sample size. This is one of the most common errors missed during data cleaning. Take the example below:

Have you ever had sex?

eversex	Freq.	Percent	Cum.
no	925	27.62	27.62
yes	**2,424**	72.38	100.00
Total	3,349	100.00	

The entire sample is asked, '*Have you ever had sex?*'. Three-fourths of the population (2,424 people) responded, '*Yes*'.

Have you had sex with a regular partner in the past 12 months?

regpart12mnth	Freq.	Percent	Cum.
no	784	32.34	32.34
yes	**1,640**	67.66	100.00
Total	**2,424**	100.00	

Only the people who reported that they had ever had sex (N=2,424) were then asked, '*Have you had sex with a regular partner in the past 12 months?*', and 1,640 people responded, Notice that the denominator of this variable is the same as the total number of people who ever had sex.

(An important note on this variable: During the cleaning stage, this variable should have in it only the people who reported ever having had sex, because those were the people who were asked the question. Later, however, you will analyze these data with a total population denominator. More information on this analysis is provided in Chapter 11: Data analysis)

Did you or your partner use a condom the last time you had sex with a regular partner?

condlastreg	Freq.	Percent	Cum.
no	1,502	91.70	95.54
yes	136	8.30	100.00
Total	**1,638**	100.00	

Next, only those people who reported a regular partner in the past 12 months (N=1,640) were asked, '*Did you or your partner use a condom the last time you had sex with a regular partner?*'. Notice that there are 1,638 responses and two people missing from the denominator, but it is still comprised of only the people with a regular partner.

It is very important that you go through your data set question by question and make sure that the denominators of each variable match the skip patterns in the questionnaire. The easiest way to do this is to cross-tab the screening question (example: *Have you ever had sex?*) and the follow-up question (example: *Have you had sex with a regular partner in the past 12 months?*). If there are people who answered negatively to the screening question, but were still asked the follow-up question due to interviewer or data entry error, they should be re-coded as missing. In the example below four people who reported that they had never had sex also reported that they had sex with a regular partner in the past 12 months. These four people should be cleaned from the denominator of the variable '*Did you have sex with a regular partner in the past 12 months*'.

	regpart12mnth		
eversex	no	**yes**	Total
no	921	**4**	925
yes	784	1,640	2,424
Total	1,705	1,644	3,349

A small number of these types of errors will not affect the overall population-level analysis. However, systematic and large-scale inaccuracies will. It is the responsibility of the investigators and analysts to review these errors and ensure that they are not biasing the analysis and the interpretation of results. Again, a strong supervising and editing team in the field as well as double data entry will go a long way to minimizing these problems.

Variables to check for correct denominators during cleaning

(from the expanded questionnaire)

Socio-demographic

1. Q110: In what sector do you earn a living?
   Denominator = Only those people who had an income from Q109
2. Q112: How long ago did you leave the country where you were born?
   Denominator = Only those people who are refugees from Q105
3. Q113: How many countries have you transited through or lived in since you left your home country, including the country where you currently live?
   Denominator = Only those people who are refugees from Q105
4. Q115: Why were you away from this place for one month or more?
   Denominator = Only those people had been away from the community where they currently live for one month or more from Q114

5. Q117: The last time you visited the refugees/host community, what was your reason?
   Denominator = Only those people who answered that they ever visited the camp or surrounding community from Q116
6. Q119: How old were you when you first married? and Q120: What is your current relationship status? and Q122: Are you currently living with a long term partner?
   Denominator = Only those people who had ever been married from Q118
7. Q121: Are you in a polygamous marriage?
   Denominator: Only those who reported they were currently married in Q120

Alcohol and drug use

8. Q125: What drugs have you taken? and Q126: Have you injected any drugs that were not prescribed by a health professional in the past 12 months?
   Denominator = Only those people who reported they had taken non-medically prescribed drugs in Q124
9. Q127: Have you used a needle or syringe to inject drugs that were not prescribed by a health professional... in the past 12 months?
   Denominator = Only those people who had even injected non-prescribed drugs in Q126

Circumcision

10. Q129: At what age were you circumcised?
    Denominator = Only those people who reported being circumcised in Q128

Military activity

11. Q133 and Q134: How long were you involved in military, paramilitary and police activities and are you currently involved?
    Denominator = Only those who were involved in these activities in Q132
12. Q135: How long ago did you leave your military, paramilitary and police activities?
    Denominator = Only those who were had left these activities in Q134

Sexual activity

13. Q202: At what age did you first have sex?
    Denominator = Only those who had ever had sex in Q201

Regular sex partners

14. Q204: Have you had a regular sex partner in the past 12 months?
    Denominator = Only those who had ever had sex in Q201
15. Q205-208: Number, nationality, age and last time condom use with regular sex partner
    Denominator = Only those who had a regular sex partner in Q204

Casual sex partners

16. Q209: Have you had a casual sex partner in the past 12 months?
    Denominator = Only those who had ever had sex in Q201
17. Q210-216: Number, nationality, age, marital status, profession, under the influence of alcohol and last time condom use with casual sex partner
    Denominator = Only those who had a casual sex partner in Q209
18. Q217: Who suggested condom use the last time you had sex with a casual partner?
    Denominator = Only those who used a condom with last casual partner in Q216
19. Q218: What was the main reason you did not use a condom the last time you had sex with a casual partner?
    Denominator = Only those who did not use a condom with last casual partner in Q216
20. Q219: In the past 12 months, how often did you use a condom with all of your casual sex partners?
    Denominator = Only those who had a casual sex partner in Q209

Transactional sex

21. Q220: Have you ever had sex in exchange for money, a gift, or a favor?
    Denominator = Only those who had ever had sex in Q201
22. Q221-222: The last time you exchanged sex was it for money, a gift or a favor? Who was it with?
    Denominator = Only those who had ever had transactional sex in Q220
23. Q223: During what period in your life did you exchange sex for money, a gift or a favor? (refugees only)
    Denominator = Only those who had ever had transactional sex in Q220 AND who were refugees in Q105
24. Q224: During what period in your life did you exchange sex for money, a gift or a favor? (nationals only)
    Denominator = Only those who had ever had transactional sex in Q220 AND who were not refugees in Q105
25. Q225: Have you ever had sex in exchange for money, a gift, or a favor in the past 12 months?
    Denominator = Only those who had ever had transactional sex in Q220
26. Q226-231: Number, type of transactional sex, partner characteristic, age, under the influence of alcohol and last time condom use with casual sex partner
    Denominator = Only those who had a transactional sex partner in the past 12 months in Q225
27. Q232: Who suggested condom use the last time you had sex in exchange for money, a gift, or a favor?
    Denominator = Only those who used a condom at last transactional sex in Q231
28. Q233: What was the main reason you did not use a condom the last time you had sex in exchange for money, a gift, or a favor?
    Denominator = Only those who did not use a condom at last transactional sex in Q231

29. Q219: In the past 12 months, how often did you use a condom with all of the people with whom you had sex in exchange for money, a gift, or a favor?
    Denominator = Only those who had a transactional sex in the past 12 months in Q225

Forced sex

30. Q236: During what period in your life were you forced to have sex? (refugees only)
    Denominator = Only those who had ever been forced to have sex in Q235 AND who were refugees in Q105
31. Q237: During what period in your life were you forced to have sex? (nationals only)
    Denominator = Only those who had ever been forced to have sex in Q235 AND who were not refugees in Q105
32. Q238: Who forced you to have sex?
    Denominator = Only those who had ever been forced to have sex in Q235
33. Q239: If you were forced by a non-family member, who forced you?
    Denominator = Only those who had been forced by a non-family member in Q238
34. Q240: Have you been forced to have sex in the past 12 months?
    Denominator = Only those who had ever been forced to have sex in Q235
35. Q241-242 and 244: How many times were you forced to have sex? By whom? How old were they?
    Denominator = Only those who had been forced in the past 12 months in Q240
36. Q243: If you were forced by a non-family member in the past 12 months, who forced you?
    Denominator = Only those who had been forced by a non-family member in the past 12 months in Q242

Anal sex

37. Q246: The last time you had anal sex with a man, did you or your partner use a condom? (women only)
    Denominator = Only those who had anal sex with in past 12 months in Q245 AND who were female in Q101
38. Q247: Have you had anal sex with a man in the past 12 months? (men only)
    Denominator = Only those who had anal sex with a man or a woman in past 12 months in Q245 AND who were male in Q101
39. Q248: The last time you had anal sex with a man, did you or your partner use a condom? (men only)
    Denominator = Only those who had anal sex with a man in past 12 months in Q247 AND who were male in Q101
40. Q249: Have you had anal sex with a woman in the past 12 months? (men only)
    Denominator = Only those who had anal sex with a man or a woman in past 12 months in Q245 AND who were male in Q101
41. Q250: The last time you had anal sex with a woman, did you or your partner use a condom? (men only)

Denominator = Only those who had anal sex with a woman in past 12 months in Q249 AND who were male in Q101

Male and female condoms

42. Q302: What do you think condoms are used for?
    Denominator = Only those who had ever heard of condoms in Q301
43. Q303: Have you ever used a condom?
    Denominator = Only those who had ever heard of condoms in Q301
44. Q304: Do you know where to obtain a condom?
    Denominator = Only those who had ever used a condom in Q303
45. Q305-306: Where do you usually get condoms? Can you get one every time you need one?
    Denominator = Only those who knew where to obtain a condom in Q304
46. Q307: What is the main constrain to obtaining a condom?
    Denominator = Only those who cannot get a condom every time they need one in Q306
47. Q308: Have you heard of a female condom?
    Denominator = Only those who had ever heard of condoms in Q301
48. Q309-311: Have you ever used a female condom? Would you be willing to use one? Do you know where to obtain one?
    Denominator = Only those who had ever heard of female condoms in Q308

Sexually transmitted infections

49. Q404: During the last time you had genital discharge, ulcer or sore, did you seek treatment?
    Denominator = Only those who had unusual genital discharge in Q402 OR a genital ulcer or sore in Q403
50. Q405: Where was the first place you went for treatment?
    Denominator = Only those who sought treatment in Q404
51. Q406: During the last time you had genital discharge, ulcer or sore, did you inform your sexual partner?
    Denominator = Only those who had unusual genital discharge in Q402 OR a genital ulcer or sore in Q403

Knowledge, opinions, and attitudes towards HIV

52. Q502: Do you think there are more cases of HIV/AIDS in your community or the surrounding local community?
    Denominator = Only those who had ever heard of HIV in Q501 AND were a refugee in Q105
53. Q503: Do you think there are more cases of HIV/AIDS in your community or the refugee community?
    Denominator = Only those who had ever heard of HIV in Q501 AND were not a refugee in Q105
54. Q504-520: All knowledge, opinions, attitudes questions in this section
    Denominator = Only those who had ever heard of HIV in Q501

Exposure and access to interventions

55. Q601: Have you received HIV/AIDS information in the past 12 months? *Denominator = Only those who had ever heard of HIV in Q501*
56. Q602: From what sources have you received information on HIV/AIDS in the past 12 months?
    Denominator = Only those who had received HIV/AIDS information in Q601
57. Q603: From what sources would you prefer to receive information on HIV/AIDS in the past 12 months?
    Denominator = Only those who had ever heard of HIV in Q501
58. Q604: Do you know a place where a person can be tested for HIV?
    Denominator = Only those who had ever heard of HIV in Q501
59. Q605: Where can a person be tested for HIV?
    Denominator = Only those who reported that that the knew of a place where a person could be tested in Q604
60. Q607-611: When was the last time you were tested for HIV? Was it mandatory or voluntary? Did you receive counseling? Where did you go? Did you obtain the results of the test?
    Denominator = Only those who had ever been tested for HIV in Q606
61. Q613: What is the primary reason you don't want to go for a test?
    Denominator = Only those who would not go for an HIV test in the future in Q612

Pregnancy

62. Q615: Have you been pregnant in the past 5 years?
    Denominator = Only those who are women in Q101
63. Q616: When you were pregnant did you go to an ante-natal clinic?
    Denominator = Only those women who reported pregnancy in Q615

Important Note: These are the denominators we are using for cleaning because they correspond to the questionnaire. However, there will be some differences between these denominators and the ones you will use to analyze the data. **Please refer to this list ONLY as a reference for data cleaning**. More information on the denominators used during analysis is provided in *Chapter 11: Data analysis* and *Appendix 5: Coding and analysis guide*

Step 5:
Use your common sense and pay attention to detail!

Most data errors can be corrected if the person responsible for cleaning the data sets is very familiar with the questionnaire and data entry program, develops a data cleaning plan, and above all is conscientious about paying attention to detail. Make sure that the person cleaning the data has the skills and training to be able to ensure that the datasets are as clean as possible data analysis begins.

A missing value is not a non-response.

If someone was asked a question and did not know the answer or did want to respond to the question, it is usually coded in the data set as '98' (no response). If a person was not asked the question at all due to a skip pattern in the question, no data should be recorded in the data set and it if considered a missing value. Never change a missing value to a non-response (98) because in doing that you are changing the denominator of the question.

An example of a common denominator error and how to clean it

A data error that most surveys encounter is one where there are inconsistent answers about sexual activity. For example, people who have reported that they didn't have a casual sex partner in the past 12 months but then enumerated the number of times they had sex with a casual sex partner in the past 12 months, reported on the nationality of their last casual partner and whether they used condoms with that partner, etc... This inconsistency may be due to data entry or interviewer error.

The variable could have been entered incorrectly by the data entry clerk. Or maybe the participant did have a casual sex partner in the past 12 months and the interviewer recorded the answer incorrectly on the questionnaire. Or maybe the participant did not have a casual sex partner in the past 12 months, but had sometime before that and was giving answers about partners outside of the time reference period because the interviewer incorrectly asked them the follow-up questions.

The later two examples are problems that should have been corrected by the interviewer and team leader when reviewing and editing the questionnaire. It is impossible to know which scenario is correct once you get to data cleaning. The rule of thumb is *do not ever change a participant's answer*, even if it may be the wrong one.

→

There is no way for the data manager to make those determinations. You can do several things to correct this error. First, you can go back to the original questionnaire to make sure the answer was entered correctly (unless you have done double data entry, and then that error would have already been corrected). If it was entered correctly, you can keep the response to the initial question '2 - *No, I did not have a casual partner in the past 12 months*' and recode the rest of the values in the section to missing. Or you can recode the entire section to missing. What you cannot do is recode the initial question to '1 - *Yes, I did have a casual partner in the past 12 months*'.

If the questionnaires were being carefully edited in the field, you should not encounter this type of interviewer error too often and the missing values should not change the results of your analysis. If you are finding this problem frequently, cross-check whether the errors are occurring with a single interviewer. If they are, this may mean that you will need to make the difficult determination about whether any of the interviewer's data is reliable. Both keeping that interviewer's questionnaires in the data sets and removing them completely could introduce bias into the survey. The data manager should consult with the field supervisor and other investigators about the proper course of action in this case and the result and possible biases should be well-documented in the report.

Renaming variables

A very useful tool for analysis is to give each of the variables in your data set a new name that corresponds directly to the question they contain. For example, you may want to analyze whether a person had a casual partner in the past 12 months by nationality. In your data set for the expanded questionnaire, the casual partner question is *Q208* and the nationality question is *Q104*. However, to find this information you would have to reference the questionnaire and coding guide which slows down the analysis.

If you give your variables new names that are immediately recognizable then you can speed up your analysis considerably and minimize the chance that the wrong variable is used. For example, *Q208* becomes *caspart12m* and *Q104* becomes *nationcurr*.

It is not necessary, or always possible, to rename your variables, but doing so may help your analysis and also make the data sets easier to share with people who are unfamiliar with your survey. In *Appendix 5: Coding and analysis guide*, recommended variable names are provided for all questions in the expanded questionnaire.

Two important notes about re-labeling:

1. Not all of the variable names in the *Coding and analysis guide* can be used in EpiInfo because the program limits variable names to 8 digits.

2. Any recoding should be done in a separate data set and you should always keep the master file with the original variable names.

Storing fieldwork forms

Always have a central place where all questionnaires are stored and develop a system such as a check-list for identifying what questionnaires have been entered. Completed questionnaires can become mixed up with those that have not been entered, and vice versa, which will cause confusion. It is also important to have the questionnaires in an accessible place during data cleaning in case they need to be referenced. The international standard is for questionnaires to be stored for five years after the survey before they should be discarded, but this many vary from location to location and is not always necessary.

Checklist 7: Data management

- ☐ Choose appropriate data entry software
- ☐ Design data entry program
- ☐ Train data entry clerks
- ☐ Develop data entry schedule
- ☐ Enter data (using double data entry whenever possible)☐ Combine data sets
- ☐ Clean data
  - o Merge data sets from double data entry
  - o Review for completeness
  - o Screen for outliers
  - o Check denominators
- ☐ Rename variables (optional)
- ☐ Store questionnaires and fieldwork forms in central location

Chapter 11

Data analysis

Tools for analysis

Just as with all the survey steps that came before, data analysis is done most efficiently if first you have a plan. The first tool for assisting your analysis is the **coding guide**. This document lays out each variable name in the questionnaire, the recode name (if applicable), the description of the variable, and the codes and description for each response option. An example section of a coding guide is below.

Table 5: Example coding guide

Variable	Rename (optional)	Description	Code
Q109	Income	Has an income	1=Yes 2=No
Q110	Incomesector	In what sector earns an income	1=Agriculture 2=Trading 3=Pastoralism 4= Transport 5=Fishing 6=Crafts 7=Private services 8=Public services 9= Humanitarian or development group 10=Other
Q110a	Othersector	Other sector where earns an income	1= Construction 2 = Business 3= Unspecified
Q111	Timeres	Length of time living in current community	1=Always 2=Less than 6 months 3=Between 6-12 months 4= 1-2 years 5= 3-5 years 6= More than 5 years 99= Don't know

As you can see from the above example, it is much easier to conduct an analysis if you have a single, user friendly guide to the different variable codes and definitions. Having to page through the questionnaire for reference can be difficult for the analyst. In addition, the questionnaire will not have the variable 'Q110a' (othersector), which enumerates the 'other' responses given to the question '*In what sector do you earn an income?*' because this variable was added when creating the database.

The coding guide is also a necessary reference for comparing data across different data sets, particularly when not all of the variable names, codes and time reference periods are the same and the analysts are unfamiliar with the questionnaires and raw data.

In order to standardize and document your analysis, it is also useful to expand this coding guide to be a **coding and analysis guide**. What this entails is thinking – ahead of time – about how you are going to analyze each variable and, in particular, considering which denominator you are going to use. As stated in *Chapter 10: Data management*, the denominators in your data sets are not always the final denominators for your analysis.

The example below shows two additions to your coding guide. First, the denominator that will be used for analysis is included in the variable description. Second, any notes on how the analysis should be done are included in the right-hand column.

Table 6: Example coding and analysis guide

Variable	Rename (optional)	Description (with denominator)	Code	Notes
Q109	Income	Has an income/ *Total population*	1=Yes 2=No	
Q110	Incomesector	In what sector earns an income/ *Had an income (Q109)*	1=Agriculture 2=Trading 3=Pastoralism 4= Transport 5=Fishing 6=Crafts 7=Private services 8=Public services 9= Humanitarian or development group 10=Other	

Variable	Rename (optional)	Description (with denominator)	Code	Notes
Q110a	Othersector	Other sector where earns an income/ *Reported other income sector (Q110)*	1= Construction 2 = Business 3= Unspecified	Recode 1 and 2 to 'private sector' in Q110.
Q111	Timeres	Length of time living in current community/ *Total population*	1=Always 2=Less than 6 months 3=Between 6-12 months 4= 1-2 years 5= 3-5 years 6= More than 5 years 99= Don't know	

Notice the responses 'construction' and 'business' in 'Q110a - other income sector', can be recoded to 'private sector' in Q110. This leaves the rest of the 'other income sector' responses as 'unspecified'. How you analyze the 'other' responses will depend on how many there are and how they have been entered into your database. See *Chapter 10: Data management* for more information on this topic.

Also note that the denominator for '*In what sector do you earn a living?*' (*Q110*) is not the total population. All denominators should be described in the analysis guide so they are not open to question.

Remember, these surveys are intended to be replicated, and this includes the analysis. They will also be used for comparison across sites. The better the coding and analysis is documented, the more easily the objectives of the surveys can be fulfilled. In order to assist you in doing this, a **coding and analysis guide** for the expanded questionnaire is provided in Appendix 5 for your reference. This document should be revised so that it fits any changes you have made to your particular questionnaire and is in a format that the analyst can easily use.

Recoding denominators

It cannot be overstated how important it is to pay close attention to the denominator you are using in your analysis. Sometimes, it will not be the same denominator that is in your data set because of the skip patterns. Take this common example.

Indicator: Proportion of people who have had a casual sex partner in the past 12 months.

Question 208 in the expanded questionnaire asks *'Have you had sex with a casual sex partner in the past 12 months?'*. None of the people who responded that they had never had sex (Q201) are included in the casual sex partner denominator because of the skip pattern. However, the analysis of this indicator is for the proportion of the total population who had a casual sex partner in the past 12 months. If a person has never had sex, then they also have not had the chance of being exposed to HIV through sex with a casual partner. In this case they should be recoded from 'missing' to 'no' and put back into the denominator in your analysis.

If you leave the people who have never been sexually active out of your denominator, then your analysis answers the question, *'What proportion of people who have ever been sexually active have had sex with a casual sex partner in the past 12 months?'*. While you may be interested to know the answer to this question, it does not tell you the total population who may be at risk of transmitting HIV through sex with a casual sex partner.

All of the denominators you will use for your initial analysis are documented in the coding and analysis guide for the expanded questionnaire. You can also do additional analysis of the data using different denominators, but you should make sure that you make it clear which question you are answering and which denominator you are using when you document your results.

Survey analysis

The type of analysis you will do is dependent on the survey design. *Chapter 5: Sampling* discusses the different sampling methodologies that can be used in your survey, and which one you choose will be based on the type of population data that is available and how your survey populations are organized.

The simple random sample is the 'gold standard' of sampling methodologies because every person in the population universe has an equal chance of being selected for the survey. If you selected the households in your survey using simple random sampling from complete household lists, then you can analyze your data using the basic analysis commands. You can use any analysis package including EpiInfo, SPSS, SAS or STATA to analyze data collected using a simple random sample.

However, if you collected the data using a multi-stage cluster sampling methodology (see *Chapter 5: Sampling* for more information), then you will always have to adjust your analysis to account for design effect, a phenomenon inherent in cluster sampling, and which tends to cause a loss in precision, i.e. wider confidence intervals. In the case of the BSS sampling in this manual, households selected though cluster sampling are located within close proxim-

ity of each other in each cluster instead of randomly dispersed throughout the entire survey area (which is what would be achieved through simple random sampling). Households within the same cluster are likely to share common characteristics and thus provide similar responses: because of this, households within clusters are not as representative of the true heterogeneity within the population as individual households selected through simple random sampling would be. In other words, households within clusters, taken individually, provide less information about the population than households in a simple random sample - essentially, it is as if the effective sample size were smaller.

The design effect is a value that expresses the price to pay for this loss in sample heterogeneity. It increases as a function of cluster size (thus, having many small clusters is better than a few big ones) and the actual similarity of responses within clusters (also called intra-cluster correlation). A design effect of 3.0 can be interpreted to mean that the same level of precision would have been obtained with one third as many households, if only random sampling had been used.

If the design effect is large, then the standard error for your survey is higher. In turn, your confidence intervals are wider, and you will find fewer statistically significant differences in your results. The design effect does not change the point estimates or the means.

Adjusting for design effect is important anytime you are presenting confidence intervals or testing for statistical significance. It is particularly important when you are measuring trends over time, as these surveys are intended to do. EpiInfo can handle simple analysis with adjustments for design effect. STATA can do a wide range of calculations, including logistic regression, accounting for design effect. SPSS 14.0 can adjust for design effect, but earlier versions of SPSS cannot and should never be used when analyzing data collected using a cluster sampling methodology.

Analyzing proportions and means

The first step in your analysis is to calculate and document all of the proportions or means for each of the variables. All analysis should be done by **sex** and **displacement status**, and in most cases **age group** (usually 15-19, 20-24, 15 to 24 and 25-49, as specified in Table 1). The 'n' and total for each variable should also be recorded. In you will find and example of a BSS report in Uganda that used the expanded questionnaire. This document provides you with examples of how this data can be analyzed and documented, though please note that some of the indicators have been revised since the time the survey took place.

In , we discussed the key survey indicators, their definition, denominator and level of disaggregation. This section will take you through the analysis of each of these indicators using the expanded questionnaire.

Indicator 1: Young men and women aged 15-24 who have had sexual intercourse before the age of 15 years

Indicator 1	Definition	Construction Using Expanded Questionnaire
Young men and women aged 15-24 who have had sexual intercourse before the age of 15 years	Percent of men and women aged 15-24 who have had sexual intercourse before the age of 15	**Numerator:** Q202=<15 years If missing (never had sex), re-code as 'no' *First sex before age 15* **Denominator:** Age=15-24 **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-19, 20-24, and 15-24

Indicator 2: Never-married young people aged 15–24 who have never had sex

Indicator 2	Definition	Construction Using Expanded Questionnaire
Never-married young people aged 15–24 who have never had sex	Percent of men and women aged 15-24 who never had sex	**Numerator:** Q201=2 (no) *Never had sex* **Denominator:** Q118=2 (no) *Never been married* AND Age=15-24 **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-19, 20-24, and 15-24

Indicator 3: More than one sex partner in the past 12 months among men and women aged 15-49

Indicator 3	Definition	Construction Using Expanded Questionnaire
More than one sex partner in the past 12 months among men and women aged 15-49	Percent of men and women aged 15-49 who report having sex with more than one regular, non-regular and/or transaction partners	**Numerator:** Q205 (*number regular sex partners*) + Q210 (*number casual sex partners*) + Q226 (*number transactional sex partners*) = >1 *Sex with more than one partner in the past 12 months* **Denominator:** Age=15-49 **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-19, 20-24, 15-24, and 25-49

Indicator 4: More than one sex partner in the past 12 months and reported using a condom during last sexual intercourse among men and women aged 15-49

Indicator 4	Definition	Construction Using Expanded Questionnaire
More than one sex partner in the past 12 months and reported using a condom during last sexual intercourse among men and women aged 15-49	Women and men aged 15-49 who had more than one sex partner in the past 12 months and reported using a condom during last sexual intercourse	**Numerator:** Q203 (*used condom at last sex*) =1 (yes) **Denominator:** Q204 (*number regular sex partners*) + Q210 (*number casual sex partners*) + Q226 (*number transactional sex partners*) = At least one response = yes *Indicator 3* **AND** Age =15-49 **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-19, 20-24, 15-24, 25-49

Indicator 5: Sex with a non-regular partner in the last 12 months among men and women aged 15-49

Indicator 5	Definition	Construction Using Expanded Questionnaire
Sex with a non-regular partner in the last 12 months among men and women aged 15-49	Percent of men and women aged 15-24 who reported having sex with a non-regular partner in the past 12 month	**Numerator:** Q209=1 (yes) *Sex with a non-regular partner in the past 12 months* **Denominator:** Age=15-49 **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 6: Condom use at last sex with a non-regular partner among men and women aged 15-49

Indicator 6	Definition	Construction Using Expanded Questionnaire
Condom use at last sex with a non-regular partner among men and women aged 15-49	Percent of men and women who say they used a condom the last time they had sex with a non-regular partner, of those who had sex with a non-regular partner in the last 12 months	**Numerator:** Q216=1 (yes) *Used condom last sex with non-regular partner* **Denominator:** Q209=1 (yes) *Had sex with non-regular partner in past 12 months* **AND** Age=15-49 **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 7: Sex with a transactional partner in the last 12 months among men and women aged 15-49

Indicator 7	Definition	Construction Using Expanded Questionnaire
Sex with a transactional partner in the last 12 months among men and women aged 15-49	Percent of men and women aged 15-49 who reported having sex with a transactional partner in the past 12 month	**Numerator:** Q225=1 (yes) *Sex with a transactional partner in the past 12 months* **Denominator:** Age=15-49 **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 8: Condom use at last sex with a transactional partner among men and women aged 15-49

Indicator 8	Definition	Construction Using Expanded Questionnaire
Condom use at last sex with a transactional partner among men and women aged 15-49	Percent of men and women who say they used a condom the last time they had sex with a transactional partner, of those who had sex with a transactional partner in the last 12 months	**Numerator:** Q231=1 (yes) *Used condom last sex with transactional partner* **Denominator:** Q225=1 (yes) *Had sex with transactional partner in past 12 months* **AND** Age=15-49 **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 9: Percent of men and women aged 15-49 who received the results of their HIV test in the past 12 months

Indicator 9	Definition	Construction Using Expanded Questionnaire
Percent of men and women aged 15-49 received an HIV test in he past 12 months and know their results	Men and women aged 15–49 who have been tested for HIV in the last 12 months and received their test results the last time they were tested	**Numerator:** Q607=1 *Had HIV test in past 12 months* **AND** Q611=1 (yes) *Received results of last test* **Denominator:** Total population **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-19, 20-24, 15-24, 25-49

Indicator 10: Percent of men and women aged 15-49 who had an STI symptom in the past 12 months and sought treatment at a health facility

Indicator 10	Definition	Construction Using Expanded Questionnaire
Percent of men and women aged 15-49 who had an STI symptom in the past 12 months and sought treatment at a health facility	Men and women aged 15–49 who have report an STI symptom (genital ulcer or sore, unusually genital discharge) in the last 12 months and went to a public or private health facility as their FIRST recourse for treatment	**Numerator:** Q405= 1 **OR** 2 *Sought STI treatment at a public or private health facility* **Denominator:** Q402=1 (yes) *Unusual genital discharge in past 12 months* **OR** Q403=1 (yes) *Genital ulcer or sore in past 12 months* **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 11: Percent of men and women aged 15-49 with comprehensive correct knowledge of HIV/AIDS

This is a composite indicator constructed from the 5 prompted knowledge and misconceptions questions. The person must respond correctly to all 5 questions.

Indicator 11	Definition	Construction Using Expanded Questionnaire
Percent of men and women aged 15-49 with comprehensive correct knowledge of HIV/AIDS	Men and women who correctly identify two major ways of preventing HIV sexual transmission: ▪ Using condoms ▪ Limiting sex to one faithful, uninfected partner AND who reject the two most common misconceptions: ▪ Mosquitoes transmit HIV ▪ Sharing food with an infected person transmits HIV AND who knows that: ▪ A healthy-looking person can transmit HIV	**Numerator:** Q505=1 (yes) *Know condoms prevent HIV* **AND** Q504=1 (yes) *Know sex with only one faithful, uninfected partner prevents HIV* **AND** Q507=2 (no) *Do not think mosquitoes transmit HIV* **AND** Q511=2 (no) *Do not think sharing food transmits HIV* **AND** Q512=1 (yes) *Knows healthy-looking person can have HIV* **Denominator:** Total population *(Note: Missing values –those people who have never heard of HIV - should be recoded as no)* **Disaggregate** by gender (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 12: Percent of men and women aged 15-49 with accepting attitudes towards PLHIV

This is a composite indicator constructed from the 4 prompted attitudes questions. The person must respond correctly to all 4 questions.

Indicator 12	Definition	Construction Using Expanded Questionnaire
Percent of men and women aged 15-49 with accepting attitudes towards PLHIV	Men and women who report that they would ▪ Be willing to care for a family member sick with AIDS in their own household AND ▪ Would be fresh vegetables from a shopkeeper with HIV AND ▪ Feel a teacher with HIV should be allowed to continue working AND ▪ Does not feel that it should be kept a secret if a family member had HIV	**Numerator:** Q516=1 (yes) *Willing to care for sick family member* **AND** Q518=1 (yes) *Willing to buy vegetables from shopkeeper with HIV* **AND** Q517=1 (yes) *Think teacher with HIV should continue working* **AND** Q515=2 (no) *Does not think a family member with HIV should remain secret* **Denominator:** Q501=1 (yes) All people who have ever heard of HIV **Disaggregate** by gender (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 13: Percent of men and women aged 15-49 who have been reached by HIV prevention programmes

Indicator 13	Definition	Construction Using Expanded Questionnaire
Percentage of men and women aged 15-49 who have been reached by HIV prevention programmes	Percent of men and women aged 15-49 who knew where the could receive and HIV test and had been given condoms in the past 12 months	**Numerator:** Q604=1 *Know where to get an HIV test* **AND** Q614=1 *Have been given condoms by prevention program in past 12 months* **Denominator:** Total population **Disaggregate** by gender (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 14: Percent of women aged 15-49 who were forced to have sex in the past 12 months

Indicator 14	Definition	Construction Using Expanded Questionnaire
Percent of women aged 15-49 who were forced to have sex in the past 12 months	Percent of women aged 15-49 who reported that they were forced to have sex in the past 12 months	**Numerator:** Q240=1 *Forced to have sex in past 12 months* **Denominator:** Total population of women Q 101=2

Indicator 15: Percent of men and women residing in current community for 12 months or less

Indicator 15	Definition	Construction Using Expanded Questionnaire
Percent of men and women residing in current community for 12 months or less	Percent of men and women aged 15-49 who reporting that they had resided in current community for 12 months or less	**Numerator:** Q111=2 **OR** 3 *Resided in community for 12 months or less* **Denominator:** Total population **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 16: Percent of men and women away from home for 4 or more weeks in the past 12 months

Indicator 16	Definition	Construction Using Expanded Questionnaire
Percent of men and women away from home for 4 or more weeks in the past 12 months	Percent of men and women aged 15-49 who report that they had been away from home for four or more weeks in the past 12 months	**Numerator:** Q114=1 *Away for 4 or more weeks in past 12 months* **Denominator:** Total population **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-24, 25-49

Indicator 17: Percent of men and women who visit the surrounding community one or more times a month

Indicator 17	Definition	Construction Using Expanded Questionnaire
Percent of men and women who visit the surrounding community one or more times a month	Percent of men and women aged 15-49 who report visiting the surrounding community one or more times a month	**Numerator:** Q116=2 **OR** 3 *Visit surrounding community one or more times a month* **Denominator:** Total population **Disaggregate** by sex (Q101, 1=male, 2=female) Age groups 15-24, 25-49

In-depth analysis

This chapter has focused on analyzing the standard proportions, means and core indicators for the survey. However, you will frequently want to do other analysis of the data sets to answer specific questions. For example, are there differences in condom use between people who have been exposed to different HIV interventions? What are the characteristics of people using health facilities to treat STIs versus those people who go to traditional practitioners? Are there differences in the proportion of women who had been forced to have sex by country of origin?

The data analyst on the survey may not know what additional questions are important to answer in order to better design and monitor HIV prevention and mitigation programs. To do a more complete analysis, it is ideal to team program managers and community members with the analysts. The former know what kind of information they need to improve programming and practices, and the later know how to get it for them from the data set. The data analyst should not be expected to know all of the questions that need answering in the community, nor should they be expected to work without input from the people who will use the data.

Analysis tables and presenting results

Data from these surveys should always be disaggregated by population (displaced or host community), sex, and survey site (where more then one area of the country is being surveyed). You should also disaggregate many of the key behavioral results by age groups (15-19, 20-24

and 25-49). The level of disaggregation is provided in the previous tables describing the calculation of each indicator.

When calculating and presenting statistical differences, confidence intervals and p values should be documented in the report tables and discussed in the body of the text.

Please contact the HIV/AIDS Unit at UNHCR (hivaids@unhcr.org) for templates for presenting the survey data.

Checklist 8: Data analysis

- ☐ Finalize the Coding and Analysis Guide
- ☐ Analyze proportions and means
- ☐ Analyze core indicators
- ☐ Fill in results templates
- ☐ Conduct other in-depth analysis
- ☐ Check your analysis for accuracy

Chapter 12

Reporting and Dissemination

What happens after the BSS data is collected, data entry and management is completed, and the analysis of the data is done? Sometimes, nothing at all. Designing and implementing Behavioral Surveillance Surveys – or any survey for that matter – is a huge expenditure of human resources and funding. Even more importantly, they involve the cooperation, participation and time of the community members and the organizations providing services to these populations. There are many people expecting to see and then use the results. However, the results are not always forthcoming. While researchers most often have the best of intentions to write up the data and present the results for the betterment of the communities, this activity often gets sidelined.

Interpreting and reporting BSS data is an important undertaking that often feels overwhelmingly complex and daunting to the implementing agencies, and it requires a breadth of knowledge and skills to be achieved successfully. The results must first be carefully interpreted by somebody who understands statistics and the survey methods (preferably the same person who did the data analysis). The findings need to be presented in a format that is clear and easy to understand, but does not sacrifice statistical documentation. The results must then be explained and contextualized in a manner that is beneficial to the end-users, who frequently are not researchers themselves. And conclusions should be developed with specific recommendations that can be acted upon.

Often this process leads to a final report (months and sometimes years after data collection), but few other mechanisms are developed for getting the results to the people and organizations that will be able to use them the most. Before undertaking a BSS, it should be clear who is in need of the data, how they are going to use it, and in what formats it will be best received and understood. No single person should be expected to have all of the necessary knowledge and skills to see this entire process through. Interpreting and using BSS results is a group activity that should involve the research team, organizations that provide HIV programs for the populations, and the community members themselves.

The following is a summary of important areas to cover in the survey report.

1. Write a brief, concise **executive summary** that provides an overview of the objectives and methods of the survey as well as results of key indicators and important recommendations and conclusions. Often, people will read only the executive summary of a report, so it should provide enough information to convey the 'flavor' of the survey methods and findings without being too long.

2. Provide an **introduction** to what is known about the **HIV epidemic** in the survey area and populations and **programs and services** that are in place to curb its impact.

3. Explain the survey **background and objectives** and the contributions of the various **partners** in the research.

4. **Describe the survey area**, the **geographic boundaries** of the survey and a detailed **population description including participant inclusion criteria**. In order to understand how the BSS findings can be applied, the users of your results must be provided with detailed information on the population and survey coverage area. Include maps of the area(s) whenever they are available.

5. Explain **sample size calculations and sampling methods**. In order to achieve the objectives of the BSS, the sampling must be conducted in a statistically rigorous manner. The report should detail not only what the sample sizes were, but how they were calculated and whether they were met. Sampling methods and how sampling was ultimately carried out in the field should be described in detail.

6. Describe how the **questionnaire** was designed, translated and tested.

7. Detail how the **survey team** was recruited, trained and supervised.

8. Explain the **community involvement** in survey design and implementation and how community **sensitization** was carried out.

9. Document the **challenges and limitations** of the survey included any **potential biases** that could affect interpreting the results.

10. Describe **survey findings** by key areas of interest including:
    i. Characteristics of respondents
    ii. Displacement, mobility
    iii. Population interactions
    iv. Alcohol and drug use
    v. Sexual behavior
    vi. Forced sex
    vii. Condom knowledge and use
    viii. HIV co-factors including STI symptoms and circumcision
    ix. Knowledge, opinions and attitudes towards HIV/AIDS
    x. Exposure and access to HIV/AIDS information and services

11. Pull together the key findings of the survey into **succinct conclusions and concrete recommendations** for how action can be taken. The conclusions and recommendations should tell the reader something that they would not know (or be able to prove) without the benefits of the survey. Often it is usefully to write develop summary conclusions and recommendations for each main area of findings as well as a final conclusion that pulls all of the information together.

12. Include a copy of the **final questionnaire** in the appendix.

BSSs done among refugees and the surrounding communities have been published on the internet http://www.unhcr.org/hivaids. It can be used as a template for how BSS data can be clearly interpreted and presented. However, you should also consider the other data use needs in your survey areas and what are the most effective means for disseminating the information to the people who can use it best.

The UNAIDS/WHO Working Group on Global HIV/AIDS/STI Surveillance has published *Guidelines for the Effective Use of Data from HIV Surveillance Systems* that describes ways in which this survey data can be used and disseminated to different audiences. The document can be found at http://www.who.int/hiv/strategic/surveillance/hivpubsurveillance/en/index.html.

Chapter 13

Research ethics

There are three fundamental principles guiding ethical research among human subjects.

1. **Respect for persons:** All persons have the capacity and right to make their own decisions without coercement, and special protection must be provided to vulnerable people. Participants must be given enough information to make an educated decision.
2. **Beneficence:** The researcher is responsible for the physical, mental and social well-being of the participant, and any potential risks to the participant must be minimized.
3. **Justice:** No group of persons should be put at risk for the well-being of another, and recruitment of study participants must be equitable.

While the BSS does not involve taking specimens nor does it present the potential for direct physical harm to participants due to their involvement, high levels of ethical standards must still be maintained at all times. The responsibility for the ethical conduct of research lies on every member of the survey team. It should also be considered during every step of the process from protocol development to decisions about how and where results should be disseminated.

Listed below are some of the important ethical considerations in designing and carrying out your survey which should be considered. This is in no way a complete list of all ethical considerations, but should serve to create a dialogue among survey staff and community about how the survey can uphold high ethical standards. Violating some of these areas are direct infringements on the fundamental principles of research ethics, and others (which are no less important) are breaches of trust for the community members and organizations that assist you in your work.

- Has the survey been reviewed and approved by the local ethical review board as well as any other applicable international ethical boards?
- Have local authorities been notified about your plans and do they support the work?
- Is the informed consent form clear and well-translated?
- Have both the benefits and potential harms of participation been explained?
- Have potential harms been minimized in every way possible?
- Has informed consent been obtained from all heads of household and participants who are being recruited and is their right to refuse participation being honored?
- Are the study teams adequately screened and trained so they uphold the confidentiality of the participants and treat all community members with respect?
- Are confidential survey documents such as maps and questionnaires secured in a locked place at the end of every day?
- Are team members recording identifiers such as names anywhere on the survey documents?

- Are participants being given adequate contact information to redress any grievance that may arise from the survey?
- Are ethical violations being monitored, reported and rectified?
- Will the results of the survey be well-utilized and benefit the populations in the manner outlined in the informed consent?

Chapter 14

List of resources

Survey methodology

Behavioral Surveillance Surveys: Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV, Family Health International, 2000
http://www.fhi.org/en/HIVAIDS/pub/guide/bssguidelines.htm

Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations: The SMART Protocol, January 2005, http://www.smartindicators.org/SMART_Protocol_01-27-05.pdf

Measuring and Interpreting Malnutrition and Mortality: A Manual, US Centers for Disease Control and Prevention and the World Food Programme, 2005
http://www.unhcr.org/publ/PUBL/45f6abc92.pdf

HIV surveillance

UNAIDS/WHO Working Group on Global HIV/AIDS/STI Surveillance, *Sampling techniques for HIV surveillance*, UNAIDS/WHO, FHI, in progress

Guidelines for HIV Surveillance: Compilation of Basic Material, UNAIDS/WHO Working Group on Global HIV/AIDS/STI Surveillance, 2004
http://www.who.int/hiv/pub/surveillance/cdrom/en/index.html

Second generation surveillance for HIV: The next decade, UNAIDS/WHO Working Group on Global HIV/AIDS/STI Surveillance, 2000
http://www.who.int/hiv/pub/surveillance/pub3/en/index.html

HIV indicators

HIV Surveys Indicators Database, MEASURE Evaluation
http://www.measuredhs.com/hivdata/

Monitoring the Declaration of Commitment on HIV/AIDS: Guidelines on Construction of Core Indicators, United Nations Joint Programme on HIV/AIDS (UNAIDS), 2005
http://www.unaids.org/DocOrder/OrderForm.aspx

Millennium Development Goals Indicators, United Nations Statistics Division, 2003
http://millenniumindicators.un.org/unsd/mdg/Host.aspx?Content=Indicators/OfficialList.htm

Indicators, Reporting Requirements and Guidelines for Focus Countries, United States Agency for International Development (USAID), The President's Emergency Plan for AIDS Relief, 2005
www.usaid.gov/zm/docs/annex_5_emergency_plan_indicators_guide_july_2005_ford_dec_23-30.pdf

Monitoring and Evaluation Toolkit: HIV/AIDS, Tuberculosis and Malaria, The Global Fund to Fight AIDS, Tuberculosis and Malaria, January 2006
http://www.theglobalfund.org/en/links_resources/library/evaluation_framework/

Sampling

Food and Nutrition Technical Assistance (FANTA): Sampling Guide
http://www.fantaproject.org/publications/sampling.shtml

Sample size calculator for cross-sectional surveys, freeware from Emory University,
http://www.sph.emory.edu/fficdckms/Sample_size_for_comparing_two_cross-sectional_surveys.html

SampleSX sample size calculator for cross-sectional surveys, freeware from Brixton Health,
http://www.brixtonhealth.com/index.html
Note: This tool does not calculate sample sizes for surveys measuring changes over time.

Magnani R, Sabin K, Saidel T, Heckathorn D, *Review of sampling hard-to-reach and hidden populations for HIV surveillance*, AIDS, 19 (supp 2): S67-S72, 2005

Data entry and management

First Things First: Guidelines on Coding and Management of Behavioural Surveillance Data, Family Health International, 2006
http://www.fhi.org/en/HIVAIDS/pub/survreports/firstthingsfirst.htm

EpiData Entry software and manual, http://www.epidata.dk/

CSPro (Census and Survey Processing System), US Census Board, software and manual,
http://www.census.gov/ipc/www/cspro/index.html

Data use

Guidelines for effective use of data from HIV surveillance systems, UNAIDS/WHO Working Group on Global HIV/AIDS/STI Surveillance,2004
http://www.unaids.org/EN/resources/epidemiology/epi_recent_publications/guidelines_effective_data_use_2004.asp.

Appendix 1

Household information sheet

Household information sheet (daily summary)

To be completed by team leader

Serial number of household	Number eligible people (15-49) in household	Number participants recruited	Number participants refused	Reason for household and participant refusal	Household absent			
					Date visit 1	Date return visit 2	Date return visit 3	Absent household recruited

Appendix 2

Participant information sheet

Participant Information Sheet (one per household)

To be completed by team leader

Serial number of household	Household member number	Age (yrs)	**Gender** 1. Male 2. Female	**Relationship to the head of household** 1. Household Head 2. Spouse 3. Son/ Daughter 4. Father/ Mother 5. Brother/ Sister 6. Other relative 7. Living in household but not a relative	Visit 1	Visit 2	Visit 3
					1 = Refusal 2 = Individual not eligible 3 = Questionnaire completed 4 = Questionnaire partly completed 5 = Household member absent 6 = Others (Specify) **(for each household member record the correct answer)**		

Appendix 3

Core questionnaire

Behavioural Surveillance Survey
Great Lakes Initiative against AIDS (GLIA)

Questionnaire serial number |___|___|___|___|

Household serial number |___|___|___|___|

IDENTIFICATION	
COUNTRY	|___|
REGION/ PROVINCE	|___|___|
CAMP/ SURROUNDING AREA (Camp = 1, Surrounding area = 2)	|___|
IDENTIFICATION NUMBER OF CAMP/ VILLAGE	|___|___|
CLUSTER NUMBER	|___|___|
URBAN/ RURAL (Urban = 1, Rural = 2)	|___|

Name and Code of Interviewer _____ |____|____|____|

CONTROL				
	CONTROL ON FIELD LEVEL	CONTROL IN CENTRAL OFFICE	DATA ENTRY CLERK 1	DATA ENTRY CLERK 2
NAME	----------|__|	----------|__|	----------|__|	----------|__|
DATE	----------	----------	----------	----------
REMARKS				

Date of interview: __/__/ day /__/__/month	Start of interview: __/__/ h /__/__/min
	End of the interview: __/__/ h /__/__/ min

HOUSEHOLD RECRUITMENT INFORMATION

If the household is present: Only one household information sheet to be completed for the head of household or his/her representative.
If the household is absent/abandoned: Only one household information sheet to be completed for household.

HEAD OF HOUSEHOLD

PARTICIPANT IS HEAD OF HOUSEHOLD OR REPRESENTATIVE OF HEAD OF HOUSEHOLD ..... |___|

1 = Yes
2 = No
3 = No head of household or representative present

NUMBER OF PEOPLE IN HOUSEHOLD

Total number of people living in household .....|___|___|
Total number of eligible people aged 15-49 living in household .....|___|___|

RESULT OF HOUSEHOLD RECRUITMENT

RESULT OF HOUSEHOLD REITT ..... |___|

1 = Head of household agreed to household participation
2 = Head of household refused household participation
Reason for refusal ……
3 = Household not eligible
4 = Household temporarily absent
Date and time of first visit .....
Date and time of second visit .....
Date and time of third visit .....
Reason for household's absence .....
5 = Household abandoned
6 = Household on extended travel
7 = Other (specify) .....

PARTICIPANT RECRUITMENT INFORMATION

To be completed for every eligible person in household including the head of household.

RESULT OF PARTICIPANT RECRUITMENT
RESULT OF PARTICIPANT RECRUITMENT ..... |___|
1 = Agreed to participate and interview was completed.
2 =Agreed to participate but interview was not completed
3 = Refused to participate
Reason for refusal ....
4 = Absent
Date and time of first visit .....
Date and time of second visit ....
Date and time of third visit .....
Reason for participant's absence .....
5 = Other (specify) .....

RELATIONSHIP TO HEAD OF HOUSEHOLD
PARTICIPANT'S RELATIONSHIP TO HEAD OF HOUSEHOLD ..... |___|
1 = Head of household
2 = Spouse
3 = Son/Daughter
4 = Father/Mother
5 = Brother/Sister
6 = Other relative
7 = Living in household but not a relative

If participant is not recruited because refuses (3), is absent (4) or for other reason, record age and sex of non-participant:

Record sex of the respondent	1 = Male 2 = Female	|___|
How old are you? **Record age in years**	Record number of years 99 = DON'T KNOW	|___|___|

CONSENT FORM

Hello Sir/ Madam,
My name is .....I am an interviewer from the Ministry of Health. We are conducting a behavioural survey in this community and requesting people to participate. This will help in developing better health services in your community, especially related to HIV/AIDS.

[Ask of the household head for household consent: Your household has been randomly selected and we wish to have permission to interview eligible members of your household. May we proceed? ___Yes ___No]

You've been selected randomly and we wish, with your permission, to interview you.

Be assured that we want to learn from your experience and all the information we collect will be used to help us fight against AIDS in your community, country and region. Some of the questions asked, are of a sensitive nature, but please note that your name will not be recorded in the questionnaire, and any details related to your privacy will be kept confidential. It will not be used in relation to registration, food distribution or any other services.

Your participation in this survey is very important and we rely on you to provide us with accurate information that will help us to develop effective activities to fight HIV spread.

The interview will take approximately __ minutes, but with your cooperation it can be done quickly.

May I have your permission to undertake this interview? Yes ☐ No ☐

If you do not want to participate, why.....

Signature of the interviewer that a verbal consent was obtained:

SECTION I: BACKGROUND CHARACTERISTICS (27 questions)

N°	QUESTIONS	ANSWERs	SKIP
A. Socio-demographic			
101.	Record sex of the respondent	1 = Male 2 = Female |___|	
102.	How old are you? **Record age in years**	Record number of years 99 = DON'T KNOW |___|___|	
103.	In which country were you born?	1 = Kenya 2 = Rwanda 3 = Uganda 4 = Somalia 5 = Congo (DRC) |___| 6 = Burundi 7 = Sudan 8 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include only locally relevant countries*	
104.	What is your current nationality?	1 = Kenyan 2 = Rwandan 3 = Ugandan 4 = Somalian 5 = Congolese (DRC) |___| 6 = Burundian 7 = Sudanese 8= Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include only locally relevant nationalities*	
105.	Are you currently a refugee?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERs	SKIP
106.	What is your religion?	1 = Catholic 2 = Protestant 3 = Moslem |___| 4 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include other locally relevant religions*	
107.	What is the highest level of schooling you have completed? **(different from a literacy program)**	0 = Have never attended school 1= Did not complete primary education 2 = Primary 3 = Secondary |___| 4 = College 5 = University 98 = No answer 99 = Don't know	
108.	How easy is it for you to read a paper written in i. Congolese Swahili? ii. Acholi? iii. Kinyarwanda? iv. Runyoro? v. Runyankole? vi. Other language? (Hold up a paper written in each language) CIRCLE ONE ANSWER FOR EACH QUESTION	1 = Easy 2 = Difficult 3 = Do not read at all 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 *Note: Responses should be revised to include only locally relevant languages*	
109.	Do you earn a monthly wage or salary?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERs	SKIP
110.	In what sector do you earn a living? *(Only one answer is possible. Record the principal income sector.)*	0 = None 1 = Agriculture 2 = Trading 3 = Pastoralism 4 = Transport |___|___| 5 = Fishing 6 = Crafts 7 = Private services 8 = Public services 9 = Humanitarian or development group 10 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include other locally relevant income sectors*	
111.	How long have you been living in the community where you currently live?	1 = Always 2 = Less than 6 months 3 = 6-11 months 4 = 1-2 years |___|___| 5 = 3-5 years 6 = More than 5 years 98 = No answer 99 = Don't know	
112.	In the last 12 months have you been away from the community where you currently live for one continuous month or more?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO** *go to* **114**
113.	Why were you away from this place for one month or more?	1 = Employment 2 = Trade 3 = Family-related 4 = Political reasons 5 = Military-related |___|___| 6 = School-related 7 = In jail 8 = Health-related 9 = Holiday 10 = Religion-related 11 = Other (specify) _____ 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERs	SKIP
114.	How often do you go to the camp/surrounding community to visit?	0 = Never 1 = Less than once a month 2 = Once a month |___| 3 = Many times in a month 98 = No answer 99 = Don't know	*If* **NEVER** *go to* **116**
115.	The last time you visited the refugees/ host community, what was your reason? *Only one answer can be recorded*	1 = Employment 2 = Trade 3 = Shopping/ Market 4 = Health care 5 = School 6 = Entertainment |___|___| 7 = Food 8 = Visit relative/friend 9 = Collect firewood 10 = Attend religious service 11 = Other (specify) _____ 98 = No answer 99 = Don't know Note: Responses should be revised to include other locally relevant reasons	
116.	Have you ever been married? (dowry or registered)	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO** *go to* **118**
117.	How old were you when you first married?	Age in years 99 = Don't Know |___|___|	
118.	What is your current relationship status?	1 = Currently married 2 = Never married 3 = Divorced/Separated |___| 4 = Widow/ Widower 98 = No answer 99 = Don't know	If not currently married go to 120
119.	Are you in a monogamous or polygamous marriage?	1 = Monogamous 2 = Polygamous |___| 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERs	SKIP
120.	Are you currently living with your spouse or another a long-term sex partner?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	
B. Alcohol and drug use			
121.	In the past 4 weeks, how often have you had drinks containing alcohol?	1 = Everyday 2 = At least once a week |___| 3 = At least once a month 4 = Never 98 = No answer 99 = Don't know	
122.	Have you taken any drugs that were not prescribed by a health professional in the past 12 months? (This can include orally, sniffing, injection, other locally common methods for using drugs) **Note: A health professional does not include traditional medical practioners**	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO** *go to* **126**
123.	What drugs have you taken?	1 = Marijuana |___| 2 = Khat/miraa |___| 3 = Heroin |___| 4 = Opium |___| 5 = Amphetamines |___| 6 = Drugs/herbs from traditional healer |___| 7 = Other (Specify) _____ |___| 98 = No answer 99 = Don't know *Note: Responses should be revised to include locally relevant drugs*	
124.	Have you injected any drugs that were not prescribed by a health professional in the past 12 months? **Note: A health professional does not include traditional medical practioners**	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO** *go to* **126**
125.	Have you used a needle or syringe to inject drugs that were not prescribed by a health professional that had already been used by another person in the past 12 months?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERs		SKIP
C. Circumcision				
126.	Some men and women have been circumcised, have you been circumcised?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	*If* **No**, *go to* **201**
127.	At what age were you circumcised?	Record number of years 99 = DON'T KNOW	|___|___|	

❧ SECTION II: SEXUAL HISTORY AND RISK BEHAVIOUR (49 questions)

N°	QUESTIONS	ANSWERS		SKIP
A. SEXUAL ACTIVITY				
201.	Have you ever had sexual intercourse? (Sexual intercourse is defined as penetrative vaginal or anal sex)	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	IF **NO**, *go to* **235**
202.	At what age did you first have sexual intercourse?	Age in years 99 = Don't know	|___|___|	
203.	The last time you had sex, did you use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	
B. REGULAR SEX PARTNERS				
204.	Have you had a regular sex partner in the past 12 months? (A regular sexual partner is defined as spouse or live-in sexual partner) Cross check: If 120 does not equal 1, then probe to make sure the definition of "regular partner" is understood	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	*If* **No** *go to* **209**
205.	How many regular partners did you have sex with in last the 12 months?	Record number 98 = No answer 99 = Don't know	|___|___|	

N°	QUESTIONS	ANSWERS	SKIP
206.	What was the nationality of your most recent regular partner?	1 = Kenyan 2 = Rwandan 3 = Ugandan 4 = Somali 5 = Congolese (DRC) |___| 6 = Burundian 7 = Sudanese 8 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include only locally relevant nationalities*	
207.	How old was your most recent regular partner?	Record age in years 99 = Don't know |___|___|	
208.	The LAST TIME you had sex with your regular partner, did you use a condom?	1 = Yes 2 = No |___|___| 99 = Don't know	
C. NON REGULAR PARTNERSHIP			
209.	Have you had sex with a casual partner in the past 12 months? **(A casual sex partner is defined as any sexual partner different from the one with whom you live or are married to and from whom you did not receive or give money, gifts or favors for sex)**	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	*If* **No** *go to* **220**
210.	How many casual partners did you have sex with in last the 12 months?	Record number 98 = No answer |___|___| 99 = Don't know	

N°	QUESTIONS	ANSWERS	SKIP
211.	What was the nationality of your most recent casual partner?	1 = Kenyan 2 = Rwandan 3 = Ugandan 4 = Somali 5 = Congolese (DRC) |___| 6 = Burundian 7 = Sudanese 8 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include only locally relevant nationalities*	
212.	How old was your most recent casual partner?	Record age in years 99 = Don't know |___|___|	
213.	What was the marital status of your most recent casual partner?	1 = Currently married 2 = Never married 3 = Divorced/Separated |___|___| 4 = Widow/ Widower 5 = Other (Specify) _____ 99 = Don't know	
214.	What was the profession of your most recent casual partner?	1 = Businessperson 2 = Trader 3 = Student 4 = Driver/ Truck driver |___|___| 5 = Housemaid 6 = Pastoralist 7 = Farmer 8 = Military, paramilitary, police 9 = Commercial sex worker 10 = Humanitarian or development worker 11 = Unemployed 12 = Other (Specify) _____ 99 = Don't know	
215.	The last time you had sex with a casual partner, had you taken any alcohol?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERS	SKIP
216.	The last time you had sex with a casual partner did you use a condom?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	If **No** *go to* **218**
217.	Who suggested using a condom the last time you had sex with a casual partner?	1 = My partner 2 = Myself |___|___| 3 = Joint decision 98 = No answer 99 = Don't know	*Go to* **219**
218.	What was the *main* reason you did not use a condom the last time you had sex with a casual partner? *Record only one answer*	1 = No condoms available 2 = Free condoms not available 3 = Too expensive 4 = Partner objected 5 = Don't like them 6 = Used other contraceptive |___|___| 7 = I trust my partner 8 = Didn't think of using one 9 = Don't know what condom is 10 = Want to have a child 11 = Religious reasons 12 = Unplanned sex 13 = Didn't think it was necessary 14 = Other (Specify) _____ 98 = No answer 99 = Don't know	
219.	In the past 12 months, how often did you use a condom with all of your casual sex partners?	1 = Every time |___|___| 2 = Frequently (more than 50% of the time) 3 = Sometimes (less than 50% of the time) 4 = Never 98 = No answer 99 = Don't know	
D. TRANSACTIONAL SEX			
220.	Have you had sex in exchange for money, a gift or a favor in the past 12 months?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	If **No** *go to* **230**

N°	QUESTIONS	ANSWERS	SKIP
221.	In the past 12 months, how many partners did you have sex with in exchange for money, a gift or a favor?	Record number |___|___| 99 = Don't know	
222.	The last time you exchanged sex, was it for money, a gift or a favor?	1 = Money 2 = Gift |___| 3 = Favor 4= More than one thing (example: Money and gift, money and favor, gift and favor) 98 = No answer 99 = Don't know	
223.	Who was the last person with whom you exchanged sex for money, a gift or a favor?	1 = Refugee 2 = Person from local community 3 = Military, paramilitary, police |___|___| 4 = Humanitarian or development worker 5 = UN peacekeeper 6 = Other (Specify) _____ 98 = No answer 99 = Don't know	
224.	How old was the last person with whom you exchanged sex for money, a gift or a favor?	Record age in years 99 = Don't know |___|___|	
225.	The last time you exchanged sex for money, a gift or a favor, had you taken any alcohol?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	
226.	The last time you exchanged sex for money, a gift or a favor, did you use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	*If* **No** *go to* **228**
227.	Who suggested using a condom the last time you exchanged sex for money, a gift or a favor?	1 = My partner 2 = Myself |___|___| 3 = Joint decision 98 = No answer 99 = Don't know	*Go to* **229**

N°	QUESTIONS	ANSWERS	SKIP
228.	What was the *main* reason you did not use a condom the last time you exchanged sex for money, a gift or a favor? *Record only one answer*	1 = No condoms available 2 = Free condoms not available 3 = Too expensive 4 = Partner objected 5 = Don't like them 6 = Used other contraceptive |___|___| 7 = I trust my partner 8 = Didn't think of using one 9 = Don't know what condom is 10 = Want to have a child 11 = Religious reasons 12 = Unplanned sex 13 = Didn't think it was necessary 14 = Other (Specify) _____ 98 = No answer 99 = Don't know	
229.	In the past 12 months, how often did you use a condom with all of the people with whom you exchanged sex for money, a gift or a favor?	1 = Every time 2 = Frequently (more than 50% of the time) 3 = Sometimes (less than 50% of the time) 4 = Never |___|___| 98 = No answer 99 = Don't know	
E. FORCED SEX			
230.	Have you been forced to have sex against your will in the past 12 months?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	*If* **No**, *go to* **234**
231.	How many times were you forced to have sex in the past 12 months?	Provide Number |___|___| 99 = Don't know	
232.	Who forced you to have sex? *More than one answer can be given. Record all answers*	1 = Regular partner |___| 2 = Family member other than regular partner |___| 3 = Non-family member |___|	*If* ***Regular partner or other family member only,*** *go to* **234**

N°	QUESTIONS	ANSWERS		SKIP
233.	If you were forced to have sex by a non-family member, who forced you? *More than one answer can be given. Record all answers*	1 = Refugee 2 = Person from local community 3 = Military, paramilitary, police 4 = Humanitarian or development worker 5 = UN peacekeeper 6 = Other (Specify) _____ 99 = Don't know	|___| |___| |___| |___| |___| |___| |___|___|	
F. ANAL SEX				
234.	Have you had anal sex with a man or a woman in the past 12 months? Anal sex included both penetrative and receptive anal intercourse	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No**, *go to* **301**
235.	Women only: The last time you had anal sex with a man, did you or your partner use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
236.	Men only: Have you had anal sex with a man in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No**, *go to* **238**
237.	Men only: The last time you had anal sex with a man, did you or your partner use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
238.	Men only: Have you had anal sex with a woman in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No**, *go to* **301**
239.	Men only: The last time that you had anal sex with a woman, did you or your partner use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	

SECTION III: MALE and FEMALE CONDOMS (11 questions)

N°	QUESTIONS	ANSWERS	SKIP
301.	Have you ever heard of condoms?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **401**
302.	What do you think condoms are used for? Unprompted question. Record all answers given.	1 = Protects against STI/HIV/AIDS |___| 2 = Prevents pregnancy |___| 3 = Family Planning |___| 4 = Other (Specify) _____ |___| 5 = Don't know |___|	
303.	Have you ever used a condom?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **308**
304.	Do you know where you can obtain a condom?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **307**
305.	Where do you usually get condoms? **Only one answer possible**	1 = Pharmacy 2 = Health facility 3 = At the market 4 = From my friends |___|___| 5 = At the shop 6 = Community health worker 7 = Other (Specify) _____ 98 = No answer 99 = Don't know	
306.	Can you obtain a condom every time you need one?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **YES**, *go to* **308**

N°	QUESTIONS	ANSWERS	SKIP
307.	What is the *main* constraint to obtaining a condom every time you need one? *Only one answer possible*	1 = Too far away (geographical access) 2 = Too expensive 3 = Places not open at convenient hours 4 = Not available |___|___| 5 = Fear of being seen 6 = Health worker's attitude 7 = Other (specify) _____ 98 = No answer 99 = Don't know	
308.	Have you ever heard of a female condom?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **401**

❧ SECTION IV: SEXUALLY TRANSMITTED INFECTIONS (6 questions)

N°	QUESTIONS	ANSWERS	SKIP
401.	Have you ever heard about diseases that can be transmitted through sexual intercourse?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	
402.	Have you had any unusual genital discharge in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	
403.	Have you had any genital ulcers or sores in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	*If* **NO** *to both* **402 AND 403**, *go to* **501**
404.	During the last time you had genital discharge, ulcer or sore, did you seek treatment?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	*If* **NO** *go to* **406**

N°	QUESTIONS	ANSWERS	SKIP
405.	Where was the FIRST place that you went for treatment? *Only one answer possible*	1 = Public health center 2 = Private health center 3 = Traditional healer/doctor/ practitioner |___| 4 = Pharmacy 5 = Friend or relative 6 = Other (specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include locally relevant location*	
406.	During the last time you had a sexually transmitted infection did you inform your sexual partner(s)?	1 = Yes, all of them 2 = Some of them, not all |___| 3 = No, none of them 98 = No answer 99 = Don't know	

❧ SECTION V: KNOWLEDGE, OPINIONS, and ATTITUDES towards HIV/AIDS (20 questions)

N°	QUESTIONS –	ANSWERS	SKIP
501.	Have you ever heard of HIV or a disease called AIDS?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **614**
502.	**Refugees only: Cross-check: 105=Yes** Do you think there are more cases of HIV/AIDS in your community or the surrounding local community?	1 = My (refugee) community 2 = Surrounding local community |___|___| 98 = No answer 99 = Don't know	
503.	**Nationals only: Cross-check: 105=No** Do you think there are more cases of HIV/AIDS in your community or the refugee community?	1 = My (surrounding local) community 2 = Refugee community |___|___| 98 = No answer 99 = Don't know	

N°	QUESTIONS –	ANSWERS		SKIP
504.	Can people protect themselves from HIV infection by staying faithful to one uninfected faithful sex partner?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
505.	Can people protect themselves from HIV infection by using a condom correctly every time they have sex?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
506.	Can people protect themselves from HIV infection by abstaining from sex?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
507.	Can people get infected with HIV through a mosquito bite?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
508.	Can people get infected with HIV by sharing a toothbrush with someone who is infected?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
509.	Can people get infected with HIV by having anal sex with a male partner and not using a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
510.	Can a person get infected by HIV by getting injected with a needle that was already used by someone else?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
511.	Can people get infected with HIV by sharing food with someone who is infected?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	

N°	QUESTIONS –	ANSWERS		SKIP
512.	Is it possible for a healthy-looking person to have HIV, the virus that causes AIDS?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
513.	If a member of your family got infected with the virus that causes AIDS, would you want it to remain a secret?	1 = Yes (keep it secret) 2 = No 98 = No answer 99 = Don't know	|___|___|	
514.	If a relative of yours became sick with the virus that causes AIDS, would you be willing to care for him in your own household?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
515.	If a teacher was infected with the virus that causes AIDS, should he/ she be allowed to continue teaching?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
516.	Would you buy fresh vegetables from a shopkeeper who was infected with the virus that causes AIDS?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
517.	Should young adolescents be taught how to use condoms?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
518.	What are the chances that you might get HIV?	1 = Good chance 2 = Moderate chance 3 = No chance 98 = No answer 99 = Don't know	|___|___|	

SECTION VI: EXPOSURE and ACCESS to INTERVENTIONS (15 questions)

N°	QUESTIONS	ANSWERS	SKIP
601.	Have you received information on HIV/AIDS in the past 12 months?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	*If* **No**, *go to* **603**
602.	From what sources have you received information on HIV/AIDS in the past 12 months? Unprompted question. Record all answers given	*Mass media* 1 = Radio |___| 2 = TV/ Video |___| 3 = Newspaper |___| 4 = Poster/pamphlet |___| *Health services* 5 = Health facility |___| 6 = VCT center |___| 7 = ANC/PMTCT center |___| *People* 8 = Community health worker |___| 9 = Friend |___| 10 = Family member |___| 11 = Person living with HIV/AIDS |___| 12 = Peer outreach worker |___| *Other places* 13 = School |___| 14 = Place of worship |___| 15 = Public meeting |___| 16 = Others (specify) _____ |___| *Note: Other locally relevant responses should be added including specific interventions being conducted in the camp/community.*	

N°	QUESTIONS	ANSWERS	SKIP
603.	From what sources would you *prefer* to receive information on HIV/AIDS? *Unprompted question. Record all answers given*	*Mass media* 1 = Radio |___| 2 = TV/ Video |___| 3 = Newspaper |___| 4 = Poster/pamphlet |___| *Health services* 5 = Health facility |___| 6 = VCT center |___| 7 = ANC/MTCT center |___| *People* 8 = Community health worker |___| 9 = Friend |___| 10 = Family member |___| 11 = Person living with HIV/AIDS |___| 12 = Peer outreach worker |___| *Other places* 13 = School |___| 14 = Place of worship |___| 15 = Public meeting |___| 16 = Others (specify) _____ |___| *Note: Other locally relevant responses should be added including specific interventions being conducted in the camp/community.*	
604.	Do you know a place where a person can be tested for HIV?	1 = Yes |___|___| 2 = No 98 = No answer 99 = Don't know	*If* **No or Don't know**, *go to* **606**
605.	Where can a person be tested for HIV?	1 = In local community 2 = In refugee camp |___|___| 3 = In both sites 98 = No answer 99 = Don't know	
606.	I do not want to know the results, but have you ever been tested for HIV? *(State that you do not want to know the result of the test)*	1 = Yes |___|___| 2 = No 98 = No answer 99 = Don't know	*If* **No**, *go to* **612**

N°	QUESTIONS	ANSWERS	SKIP
607.	When was the last time you were tested for HIV?	1 = In the past 12 months 2 = 1-2 years ago 3 = 3 or more years ago |___|___| 98 = No answer 99 = Don't know	
608.	The last time you were tested for HIV did you yourself ask for the test, was it offered to you and you accepted, or was it required?	1 = I asked for the test 2 = It was offered and I accepted |___|___| 3 = It was required 98 = No answer 99 = Don't know	
609.	The last time you were tested for HIV did you receive counselling?	1 = Yes |___|___| 2 = No 98 = No answer 99 = Don't know	
610.	The last time you were tested for HIV, where did you go to get tested? *Only one answer possible*.	*Public sector* 1 = Hospital |___|___| 2 = Health facility government 3 = Clinic/ family planning 4 = Mobile Clinic *Private Sector* 5 = Private hospital/ Clinic 6 = Pharmacy 7 = Private medical doctor 8 = Mobile clinic 9 = Traditional healer 10 = Other (Specify) _____	
611.	I do not want to know the result, but, the last time you were tested for HIV did you obtain the result of the test? *(State again that you do not want to know the test result)*	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	
612.	Would you go for an HIV test in the future?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	*If* **Yes**, *go to* **614**

N°	QUESTIONS	ANSWERS	SKIP
613.	What is the *primary* reason you don't want to go for a test? *Only one answer possible*	1 = Don't know where to go for a test 2 = Sure of not being infected 3 = Afraid of the result 4 = Afraid of the blood taking |___|___| 5 = (Afraid of) catching an infection 6 = Fear of stigmatisation 7 = Don't think testing is confidential 8 = Too expensive 9 = Other (Specify) _____ 98 = No answer 99 = Don't know	
614.	Have you been given condoms by an HIV prevention program in the past 12 months?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	
615.	Women only Have you been pregnant in the past 5 years?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **No, *end interview***
616.	Women only When you were pregnant did you go to an antenatal clinic?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	

That is the end of the questionnaire. Thank you for taking the time to answer our questions. We appreciate your help.

Appendix 4

Expanded questionnaire

Behavioural Surveillance Survey
Great Lakes Initiative against AIDS (GLIA)

Questionnaire serial number |___|___|___|___|

Household serial number |___|___|___|___|

IDENTIFICATION	
COUNTRY	|___|
REGION/ PROVINCE	|___|___|
CAMP/ SURROUNDING AREA (Camp = 1, Surrounding area = 2)	|___|
IDENTIFICATION NUMBER OF CAMP/ VILLAGE	|___|___|
CLUSTER NUMBER	|___|___|
URBAN/ RURAL (Urban = 1, Rural = 2)	|___|

Name and Code of Interviewer _____ |____|____|____|

CONTROL				
	CONTROL ON FIELD LEVEL	CONTROL IN CENTRAL OFFICE	DATA ENTRY CLERK 1	DATA ENTRY CLERK 2
NAME	----------|__|	----------|__|	----------|__|	----------|__|
DATE	----------	----------	----------	----------
REMARKS				

HOUSEHOLD RECRUITMENT INFORMATION

If the household is present: Only one household information sheet to be completed for the head of household or his/her representative.
If the household is absent/abandoned: Only one household information sheet to be completed for household.

HEAD OF HOUSEHOLD

PARTICIPANT IS HEAD OF HOUSEHOLD OR REPRESENTATIVE OF HEAD OF HOUSEHOLD .....
.....|___|

1 = Yes
2 = No
3 = No head of household or representative present

NUMBER OF PEOPLE IN HOUSEHOLD

Total number of people living in household .....|___|___|
Total number of eligible people aged 15-49 living in household .....|___|___|

RESULT OF HOUSEHOLD RECRUITMENT

RESULT OF HOUSEHOLD REITT ..... |___|

1 = Head of household agreed to household participation
2 = Head of household refused household participation
Reason for refusal ……
3 = Household not eligible
4 = Household temporarily absent
Date and time of first visit .....
Date and time of second visit .....
Date and time of third visit .....
Reason for household's absence .....
5 = Household abandoned
6 = Household on extended travel
7 = Other (specify) .....

PARTICIPANT RECRUITMENT INFORMATION

To be completed for every eligible person in household including the head of household.

RESULT OF PARTICIPANT RECRUITMENT
RESULT OF PARTICIPANT RECRUITMENT ..... |___|
1 = Agreed to participate and interview was completed.
2 =Agreed to participate but interview was not completed
3 = Refused to participate
Reason for refusal ....
4 = Absent
Date and time of first visit .....
Date and time of second visit ....
Date and time of third visit .....
Reason for participant's absence .....
5 = Other (specify) .....

RELATIONSHIP TO HEAD OF HOUSEHOLD
PARTICIPANT'S RELATIONSHIP TO HEAD OF HOUSEHOLD ..... |___|
1 = Head of household
2 = Spouse
3 = Son/Daughter
4 = Father/Mother
5 = Brother/Sister
6 = Other relative
7 = Living in household but not a relative

If participant is not recruited because refuses (3), is absent (4) or for other reason, record age and sex of non-participant:

Record sex of the respondent	1 = Male 2 = Female	|___|
How old are you? **Record age in years**	Record number of years 99 = DON'T KNOW	|___|___|

CONSENT FORM

Hello Sir/ Madam,
My name is .....I am an interviewer from the Ministry of Health. We are conducting a behavioural survey in this community and requesting people to participate. This will help in developing better health services in your community, especially related to HIV/AIDS.

[Ask of the household head for household consent: Your household has been randomly selected and we wish to have permission to interview eligible members of your household. May we proceed? ___Yes ___No]

You've been selected randomly and we wish, with your permission, to interview you.

Be assured that we want to learn from your experience and all the information we collect will be used to help us fight against AIDS in your community, country and region. Some of the questions asked, are of a sensitive nature, but please note that your name will not be recorded in the questionnaire, and any details related to your privacy will be kept confidential. It will not be used in relation to registration, food distribution or any other services.

Your participation in this survey is very important and we rely on you to provide us with accurate information that will help us to develop effective activities to fight HIV spread.

The interview will take approximately __ minutes, but with your cooperation it can be done quickly.

May I have your permission to undertake this interview? Yes ☐ No ☐

If you do not want to participate, why.....

Signature of the interviewer that a verbal consent was obtained:

SECTION I: BACKGROUND CHARACTERISTICS (35 questions)

N°	QUESTIONS	ANSWERs	SKIP
A. Socio-demographic			
101.	Record sex of the respondent	1 = Male 2 = Female |___|	
102.	How old are you? **Record age in years**	Record number of years 99 = DON'T KNOW |___|___|	
103.	In which country were you born?	1 = Kenya 2 = Rwanda 3 = Uganda 4 = Somalia 5 = Congo (DRC) |___| 6 = Burundi 7 = Sudan 8 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include only locally relevant countries*	
104.	What is your current nationality?	1 = Kenyan 2 = Rwandan 3 = Ugandan 4 = Somalian 5 = Congolese (DRC) |___| 6 = Burundian 7 = Sudanese 8= Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include only locally relevant nationalities*	
105.	Are you currently a refugee?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERs	SKIP
106.	What is your religion?	1 = Catholic 2 = Protestant 3 = Moslem |___| 4 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include other locally relevant religions*	
107.	What is the highest level of schooling you have completed? **(different from a literacy program)**	0 = Have never attended school 1= Did not complete primary education 2 = Primary 3 = Secondary |___| 4 = College 5 = University 98 = No answer 99 = Don't know	
108.	How easy is it for you to read a paper written in i. Congolese Swahili? ii. Acholi? iii. Kinyarwanda? iv. Runyoro? v. Runyankole? vi. Other language? *(Hold up a paper written in each language)* CIRCLE ONE ANSWER FOR EACH QUESTION	1 = Easy 2 = Difficult 3 = Do not read at all 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 *Note: Responses should be revised to include only locally relevant languages*	
109.	Do you earn a monthly wage or salary?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERs	SKIP
110.	In what sector do you earn a living? *(Only one answer is possible. Record the principal income sector.)*	0 = None 1 = Agriculture 2 = Trading 3 = Pastoralism 4 = Transport |___|___| 5 = Fishing 6 = Crafts 7 = Private services 8 = Public services 9 = Humanitarian or development group 10 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include other locally relevant income sectors*	
111.	How long have you been living in the community where you currently live?	1 = Always 2 = Less than 6 months 3 = 6-11 months 4 = 1-2 years |___|___| 5 = 3-5 years 6 = More than 5 years 98 = No answer 99 = Don't know	
112.	**Refugees only: Cross-check 105 =Yes** How long ago did you leave the country where you were born?	Record number of years |___|___| 99 = UNKNOWN	
113.	**Refugees only: Cross-check 105 =Yes** How many countries have you transited through or lived in since you left your home country, including the country where you currently live?	Record number of countries |___|___| 99 = UNKNOWN	
114.	In the last 12 months have you been away from the community where you currently live for one continuous month or more?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO** *go to* **116**

N°	QUESTIONS	ANSWERs	SKIP
115.	Why were you away from this place for one month or more?	1 = Employment 2 = Trade 3 = Family-related 4 = Political reasons 5 = Military-related |___|___| 6 = School-related 7 = In jail 8 = Health-related 9 = Holiday 10 = Religion-related 11 = Other (specify) _____ 98 = No answer 99 = Don't know	
116.	How often do you go to the camp/surrounding community to visit?	0 = Never 1 = Less than once a month 2 = Once a month |___| 3 = Many times in a month 98 = No answer 99 = Don't know	*If* **NEVER** *go to* **118**
117.	The last time you visited the refugees/ host community, what was your reason? *Only one answer can be recorded*	1 = Employment 2 = Trade 3 = Shopping/ Market 4 = Health care 5 = School 6 = Entertainment |___|___| 7 = Food 8 = Visit relative/friend 9 = Collect firewood 10 = Attend religious service 11 = Other (specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include other locally relevant reasons*	
118.	Have you ever been married? (dowry or registered)	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO** *go to* **120**

N°	QUESTIONS	ANSWERs		SKIP
119.	How old were you when you first married?	Age in years 99 = Don't Know	|___|___|	
120.	What is your current relationship status?	1 = Currently married 2 = Never married 3 = Divorced/Separated 4 = Widow/ Widower 98 = No answer 99 = Don't know	|___|	If not currently married go to 122
121.	Are you in a monogamous or polygamous marriage?	1 = Monogamous 2 = Polygamous 98 = No answer 99 = Don't know	|___|	
122.	Are you currently living with your spouse or another a long-term sex partner?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	
B. Alcohol and drug use				
123.	In the past 4 weeks, how often have you had drinks containing alcohol?	1 = Everyday 2 = At least once a week 3 = At least once a month 4 = Never 98 = No answer 99 = Don't know	|___|	
124.	Have you taken any drugs that were not prescribed by a health professional in the past 12 months? (This can include orally, sniffing, injection, other locally common methods for using drugs) **Note: A health professional does not include traditional medical practioners**	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	*If* ***NO*** *go to* ***128***

N°	QUESTIONS	ANSWERs	SKIP
125.	What drugs have you taken?	1 = Marijuana |___| 2 = Khat/miraa |___| 3 = Heroin |___| 4 = Opium |___| 5 = Amphetamines |___| 6 = Drugs/herbs from traditional healer |___| 7 = Other (Specify)_____ |___| 98 = No answer 99 = Don't know *Note: Responses should be revised to include locally relevant drugs*	
126.	Have you injected any drugs that were not prescribed by a health professional in the past 12 months? **Note: A health professional does not include traditional medical practioners**	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* ***NO*** *go to* ***128***
127.	Have you used a needle or syringe to inject drugs that were not prescribed by a health professional that had already been used by another person in the past 12 months?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	
C. Circumcision			
128.	Some men and women have been circumcised, have you been circumcised?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* ***No****, go to* ***130***
129.	At what age were you circumcised?	Record number of years 99 = DON'T KNOW |___|___|	
130.	If you could choose, would you prefer a sexual partner who was circumcised or not circumcised?	1 = Circumcised 2 = Not circumcised |___| 3 = Don't know/ no preference 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERs		SKIP
131.	**MEN ONLY** Would you be interested in getting circumcised if it was affordable and safe?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
D. Military Activity				
132.	Have you ever been involved in any official or unofficial military, paramilitary or police activities?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	*If* **NO** *go to* **201**
133.	For how long were you involved in military, paramilitary or police activities?	1 = Less than 6 months 2 = 6 to 12 months 3 = 1 to 2 years 4 = 3 to 4 years 5 = 5 or more years 98 = No answer 99 = Don't know	|___|	
134.	Are you currently involved in military, paramilitary or police activities?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	*If* **YES** *go to* **201**
135.	How long ago did you leave your military, paramilitary or police activities?	Record number of years If less than one year, record oo 99 = Don't know	|___|___|	

❧ SECTION II: SEXUAL HISTORY AND RISK BEHAVIOUR (49 questions)

N°	QUESTIONS	ANSWERS		SKIP
A. SEXUAL ACTIVITY				
201	Have you ever had sexual intercourse? (Sexual intercourse is defined as penetrative vaginal or anal sex)	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	IF **NO**, *go to* **235**

N°	QUESTIONS	ANSWERS		SKIP
202	At what age did you first have sexual intercourse?	Age in years 99 = Don't know	|___|___|	
203	The last time you had sex, did you use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	
B. REGULAR SEX PARTNERS				
204	Have you had a regular sex partner in the past 12 months? (A regular sexual partner is defined as spouse or live-in sexual partner) Cross check: If 120 does not equal 1, then probe to make sure the definition of "regular partner" is understood	1 = Yes 2 = No	|___|	*If* **No** *go to* **209**
205	How many regular partners did you have sex with in last the 12 months?	Record number 98 = No answer 99 = Don't know	|___|___|	
206	What was the nationality of your most recent regular partner?	1 = Kenyan 2 = Rwandan 3 = Ugandan 4 = Somali 5 = Congolese (DRC) 6 = Burundian 7 = Sudanese 8 = Other (Specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include only locally relevant nationalities*	|___|	
207	How old was your most recent regular partner?	Record age in years 98 = No answer 99 = Don't know	|___|___|	

N°	QUESTIONS	ANSWERS	SKIP
208	The LAST TIME you had sex with your regular partner, did you use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	
C. NON REGULAR PARTNERSHIP			
209	Have you had sex with a casual partner in the past 12 months? **(A casual sex partner is defined as any sexual partner different from the one with whom you live or are married to and from whom you did not receive or give money, gifts or favors for sex)**	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	*If* **No** *go to* **220**
210	How many casual partners did you have sex with in last the 12 months?	Record number 98 = No answer 99 = Don't know |___|___|	
211	What was the nationality of your most recent casual partner?	1 = Kenyan 2 = Rwandan 3 = Ugandan 4 = Somali 5 = Congolese (DRC) 6 = Burundian 7 = Sudanese 8 = Other (Specify) _____ 98 = No answer 99 = Don't know |___| *Note: Responses should be revised to include only locally relevant nationalities*	
212	How old was your most recent casual partner?	Record age in years 99 = Don't know |___|___|	
213	What was the marital status of your most recent casual partner?	1 = Currently married 2 = Never married 3 = Divorced/Separated 4 = Widow/ Widower 5 = Other (Specify) _____ 98 = No answer 99 = Don't know |___|___|	

N°	QUESTIONS	ANSWERS	SKIP
214	What was the profession of your most recent casual partner?	1 = Businessperson 2 = Trader 3 = Student 4 = Driver/ Truck driver |___|___| 5 = Housemaid 6 = Pastoralist 7 = Farmer 8 = Military, paramilitary, police 9 = Commercial sex worker 10 = Humanitarian or development worker 11 = Unemployed 12 = Other (Specify) _____ 98 = No answer 99 = Don't know	
215	The last time you had sex with a casual partner, had you taken any alcohol?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	
216	The last time you had sex with a casual partner did you use a condom?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	*If* **No** *go to* **218**
217	Who suggested using a condom the last time you had sex with a casual partner?	1 = My partner 2 = Myself |___|___| 3 = Joint decision 98 = No answer 99 = Don't know	*Go to* **219**

N°	QUESTIONS	ANSWERS	SKIP
218	What was the *main* reason you did not use a condom the last time you had sex with a casual partner? *Record only one answer*	1 = No condoms available 2 = Free condoms not available 3 = Too expensive 4 = Partner objected 5 = Don't like them 6 = Used other contraceptive |___|___| 7 = I trust my partner 8 = Didn't think of using one 9 = Don't know what condom is 10 = Want to have a child 11 = Religious reasons 12 = Unplanned sex 13 = Didn't think it was necessary 14 = Other (Specify) _____ 98 = No answer 99 = Don't know	
219	In the past 12 months, how often did you use a condom with all of your casual sex partners?	1 = Every time |___|___| 2 = Frequently (more than 50% of the time) 3 = Sometimes (less than 50% of the time) 4 = Never 98 = No answer 99 = Don't know	
D. TRANSACTIONAL SEX			
220	Have you ever had sex in exchange for money, a gift or a favor?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **No** *go to* **235**
221	The last time you exchanged sex, was it for money, a gift or a favor?	1 = Money 2 = Gift |___| 3 = Favor 4= More than one thing (eg: Money and gift, money and favor, gift and favor) 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERS	SKIP
222	Who was the last person with whom you exchanged sex for money, a gift or a favor?	1 = Refugee 2 = Person from local community 3 = Military, paramilitary, police |___|___| 4 = Humanitarian or development worker 5 = Other (Specify) _____ 98 = No answer 99 = Don't know	
223	**Refugees only : Cross-check 105 =Yes** During which period in your life did you exchange sex for money, a gift or a favor? *Record all answers*	**A. Before displacement** |___| 1 = Yes 2 = No **B. = During displacement** |___| 1 = Yes 2 = No **C. = After displacement** |___| 1 = Yes 2 = No	
224	**Nationals only: Cross-check 105=No** During which period in your life did you exchange sex for money, a gift or a favor? *Record all answers*	A. = Before refugees arrived |___| 1 = Yes 2 = No B. = After refugees arrived |___| 1 = Yes 2 = No	
225	Have you had sex in exchange for money, a gift or a favor in the past 12 months?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **No** *go to* **235**
226	In the past 12 months, how many partners did you have sex with in exchange for money, a gift or a favor?	Record number |___|___| 99 = Don't know	
227	The last time you exchanged sex, was it for money, a gift or a favor?	1 = Money 2 = Gift |___| 3 = Favor 4= More than one thing (example: Money and gift, money and favor, gift and favor) 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERS	SKIP
228	Who was the last person with whom you exchanged sex for money, a gift or a favor?	1 = Refugee 2 = Person from local community 3 = Military, paramilitary, police |___|___| 4 = Humanitarian or development worker 5 = UN peacekeeper 6 = Other (Specify) _____ 98 = No answer 99 = Don't know	
229	How old was the last person with whom you exchanged sex for money, a gift or a favor?	Record age in years 98 = No answer 99 = Don't know |___|___|	
230	The last time you exchanged sex for money, a gift or a favor, had you taken any alcohol?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	
231	The last time you exchanged sex for money, a gift or a favor, did you use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know |___|___|	*If* **No** *go to* **233**
232	Who suggested using a condom the last time you exchanged sex for money, a gift or a favor?	1 = My partner 2 = Myself |___|___| 3 = Joint decision 98 = No answer 99 = Don't know	*Go to* **234**
233	What was the *main* reason you did not use a condom the last time you exchanged sex for money, a gift or a favor? *Record only one answer*	1 = No condoms available 2 = Free condoms not available 3 = Too expensive 4 = Partner objected 5 = Don't like them 6 = Used other contraceptive |___|___| 7 = I trust my partner 8 = Didn't think of using one 9 = Don't know what condom is 10 = Want to have a child 11 = Religious reasons 12 = Unplanned sex 13 = Didn't think it was necessary 14 = Other (Specify) _____ 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERS	SKIP
234	In the past 12 months, how often did you use a condom with all of the people with whom you exchanged sex for money, a gift or a favor?	1 = Every time 2 = Frequently (more than 50% of the time) 3 = Sometimes (less than 50% of the time) 4 = Never |___|___| 98 = No answer 99 = Don't know	
E. FORCED SEX			
235	Have you ever been forced to have sex against your will?	1 = Yes |___| 2 = No 98 = No answer 99 = Don't know	*If* **No**, *go to* **245**
236	**REFUGEE ONLY : Cross-check 105 =Yes** During which period in your life were you forced to have sex? *Record all answers*	**A. Before displacement** |___| 1 = Yes 2 = No **B. = During displacement** |___| 1 = Yes 2 = No **C. = After displacement** |___| 1 = Yes 2 = No	
237	**Nationals only: Cross-check 105=No** During which period in your life were you forced to have sex? *Record all answers*	A. = Before refugees arrived |___| 1 = Yes 2 = No B. = After refugees arrived |___| 1 = Yes 2 = No	
238	Who forced you to have sex? *More than one answer can be given. Record all answers*	1 = Regular partner |___| 2 = Family member other than regular partner |___| 3 = Non-family member |___|	*If* ***Regular partner or other family member (1 or 2) only,*** *go to* **240**

N°	QUESTIONS	ANSWERS	SKIP
239	If you were forced to have sex by a non-family member, who forced you? *More than one answer can be given. Record all answers*	1 = Refugee |___| 2 = Person from local community |___| 3 = Military, paramilitary, police |___| 4 = Humanitarian or development worker |___| 5 = UN peacekeeper |___| 6 = Other (Specify) _____ |___| 99 = Don't know |___|___|	
240.	Have you been forced to have sex against your will in the past 12 months?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	*If* **No**, *go to* **245**
241.	How many times were you forced to have sex in the past 12 months?	Provide Number |___|___| 99 = Don't know	
242.	Who forced you to have sex? *More than one answer can be given. Record all answers*	1 = Regular partner |___| 2 = Family member other than regular partner |___| 3 = Non-family member |___|	*If* ***Regular partner or other family member only,*** *go to* **244**
243.	If you were forced to have sex by a non-family member, who forced you? *More than one answer can be given. Record all answers*	1 = Refugee |___| 2 = Person from local community |___| 3 = Military, paramilitary, police |___| 4 = Humanitarian or development worker |___| 5 = UN peacekeeper |___| 6 = Other (Specify) _____ |___| 99 = Don't know |___|___|	
244.	How old was the last person who forced you to have sex?	1 = Older than me 2 = Younger than me |___|___| 3 = Same age as me 98 = No answer 99 = Don't know	

N°	QUESTIONS	ANSWERS		SKIP
F. ANAL SEX				
245.	Have you had anal sex with a man or a woman in the past 12 months? Anal sex included both penetrative and receptive anal intercourse	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No**, *go to* **301**
246.	Women only: The last time you had anal sex with a man, did you or your partner use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
247.	Men only: Have you had anal sex with a man in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No**, *go to* **249**
248.	Men only: The last time you had anal sex with a man, did you or your partner use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
249.	Men only: Have you had anal sex with a woman in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No**, *go to* **301**
250.	Men only: The last time that you had anal sex with a woman, did you or your partner use a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	

SECTION III: MALE and FEMALE CONDOMS (11 questions)

N°	QUESTIONS	ANSWERS		SKIP
301.	Have you ever heard of condoms?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	*If* **NO**, *go to* **401**

N°	QUESTIONS	ANSWERS	SKIP
302.	What do you think condoms are used for? Unprompted question. Record all answers given.	1 = Protects against STI/HIV/AIDS |___| 2 = Prevents pregnancy |___| 3 = Family Planning |___| 4 = Other (Specify) _____ |___| 5 = Don't know |___|	
303.	Have you ever used a condom?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **308**
304.	Do you know where you can obtain a condom?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **307**
305.	Where do you usually get condoms? **Only one answer possible**	1 = Pharmacy 2 = Health facility 3 = At the market 4 = From my friends |___|___| 5 = At the shop 6 = Community health worker 7 = Other (Specify) _____ 98 = No answer 99 = Don't know	
306.	Can you obtain a condom every time you need one?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **YES**, *go to* **308**
307.	What is the *main* constraint to obtaining a condom every time you need one? *Only one answer possible*	1 = Too far away (geographical access) 2 = Too expensive 3 = Places not open at convenient hours 4 = Not available |___|___| 5 = Fear of being seen 6 = Health worker's attitude 7 = Other (specify) _____ 98 = No answer 99 = Don't know	
308.	Have you ever heard of a female condom?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **401**

N°	QUESTIONS	ANSWERS		SKIP
309.	Have you ever used a female condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	
310.	Would you/your partner be willing to use a female condom if available?	1 = Yes 2 = No 98 = No answer 99 = Don't know 99 = Don't know	|___|___|	
311.	Do you know where you can obtain a female condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|	

SECTION IV: SEXUALLY TRANSMITTED INFECTIONS (6 questions)

N°	QUESTIONS	ANSWERS		SKIP
401.	Have you ever heard about diseases that can be transmitted through sexual intercourse?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
402.	Have you had any unusual genital discharge in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
403.	Have you had any genital ulcers or sores in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **NO** *to both* **402 AND 403**, *go to* **501**
404.	During the last time you had genital discharge, ulcer or sore, did you seek treatment?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **NO** *go to* **406**

N°	QUESTIONS	ANSWERS	SKIP
405.	Where was the FIRST place that you went for treatment? *Only one answer possible*	1 = Public health center 2 = Private health center 3 = Traditional healer/doctor/ practitioner |___| 4 = Pharmacy 5 = Friend or relative 6 = Other (specify) _____ 98 = No answer 99 = Don't know *Note: Responses should be revised to include locally relevant location*	
406.	During the last time you had a sexually transmitted infection did you inform your sexual partner(s)?	1 = Yes, all of them 2 = Some of them, not all |___| 3 = No, none of them 98 = No answer 99 = Don't know	

❧ SECTION V: KNOWLEDGE, OPINIONS, and ATTITUDES towards HIV/AIDS (20 questions)

N°	QUESTIONS –	ANSWERS	SKIP
501.	Have you ever heard of HIV or a disease called AIDS?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **NO**, *go to* **614**
502.	**Refugees only: Cross-check: 105=Yes** Do you think there are more cases of HIV/AIDS in your community or the surrounding local community?	1 = My (refugee) community 2 = Surrounding local community |___|___| 98 = No answer 99 = Don't know	
503.	**Nationals only: Cross-check: 105=No** Do you think there are more cases of HIV/AIDS in your community or the refugee community?	1 = My (surrounding local) community 2 = Refugee community |___|___| 98 = No answer 99 = Don't know	

N°	QUESTIONS –	ANSWERS		SKIP
504.	Can people protect themselves from HIV infection by staying faithful to one uninfected faithful sex partner?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
505.	Can people protect themselves from HIV infection by using a condom correctly every time they have sex?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
506.	Can people protect themselves from HIV infection by abstaining from sex?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
507.	Can people get infected with HIV through a mosquito bite?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
508.	Can people get infected with HIV by sharing a toothbrush with someone who is infected?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
509.	Can people get infected with HIV by having anal sex with a male partner and not using a condom?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
510.	Can a person get infected by HIV by getting injected with a needle that was already used by someone else?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
511.	Can people get infected with HIV by sharing food with someone who is infected?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
512.	Is it possible for a healthy-looking person to have HIV, the virus that causes AIDS?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	

N°	QUESTIONS –	ANSWERS		SKIP
513.	Can a pregnant woman with HIV/AIDS, transmit the virus to her unborn child during pregnancy or delivery?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
514.	Can a woman with HIV/AIDS transmit the virus to her baby during breastfeeding?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
515.	If a member of your family got infected with the virus that causes AIDS, would you want it to remain a secret?	1 = Yes (keep it secret) 2 = No 98 = No answer 99 = Don't know	|___|___|	
516.	If a relative of yours became sick with the virus that causes AIDS, would you be willing to care for him in your own household?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
517.	If a teacher was infected with the virus that causes AIDS, should he/ she be allowed to continue teaching?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
518.	Would you buy fresh vegetables from a shopkeeper who was infected with the virus that causes AIDS?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
519.	Should young adolescents be taught how to use condoms?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	
520.	What are the chances that you might get HIV?	1 = Good chance 2 = Moderate chance 3 = No chance 4 = Already infected with HIV 98 = No answer 99 = Don't know	|___|___|	

SECTION VI: EXPOSURE and ACCESS to INTERVENTIONS (15 questions)

N°	QUESTIONS	ANSWERS		SKIP
601.	Have you received information on HIV/AIDS in the past 12 months?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No**, *go to* **603**
602.	From what sources have you received information on HIV/AIDS in the past 12 months? Unprompted question. Record all answers given	*Mass media* 1 = Radio 2 = TV/ Video 3 = Newspaper 4 = Poster/pamphlet *Health services* 5 = Health facility 6 = VCT center 7 = ANC/PMTCT center *People* 8 = Community health worker 9 = Friend 10 = Family member 11 = Person living with HIV/AIDS 12 = Peer outreach worker *Other places* 13 = School 14 = Place of worship 15 = Public meeting 16 = Others (specify) _____ *Note: Other locally relevant responses should be added including specific interventions being conducted in the camp/community.*	|___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___|	

N°	QUESTIONS	ANSWERS		SKIP
603.	From what sources would you *prefer* to receive information on HIV/AIDS? *Unprompted question. Record all answers given*	*Mass media* 1 = Radio 2 = TV/ Video 3 = Newspaper 4 = Poster/pamphlet *Health services* 5 = Health facility 6 = VCT center 7 = ANC/MTCT center *People* 8 = Community health worker 9 = Friend 10 = Family member 11 = Person living with HIV/AIDS 12 = Peer outreach worker *Other places* 13 = School 14 = Place of worship 15 = Public meeting 16 = Others (specify) _____ *Note: Other locally relevant responses should be added including specific interventions being conducted in the camp/community.*	|___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___| |___|	
604.	Do you know a place where a person can be tested for HIV?	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No or Don't know**, *go to* **606**
605.	Where can a person be tested for HIV?	1 = In local community 2 = In refugee camp 3 = In both sites 98 = No answer 99 = Don't know	|___|___|	
606	I do not want to know the results, but have you ever been tested for HIV? *(State that you do not want to know the result of the test)*	1 = Yes 2 = No 98 = No answer 99 = Don't know	|___|___|	*If* **No**, *go to* **612**

N°	QUESTIONS	ANSWERS	SKIP
607	When was the last time you were tested for HIV?	1 = In the past 12 months 2 = 1-2 years ago 3 = 3 or more years ago |___|___| 98 = No answer 99 = Don't know	
608	The last time you were tested for HIV did you yourself ask for the test, was it offered to you and you accepted, or was it required?	1 = I asked for the test 2 = It was offered and I accepted |___|___| 3 = It was required 98 = No answer 99 = Don't know	
609	The last time you were tested for HIV did you receive counselling?	1 = Yes |___|___| 2 = No 98 = No answer 99 = Don't know	
610	The last time you were tested for HIV, where did you go to get tested? *Only one answer possible*.	*Public sector* 1 = Hospital |___|___| 2 = Health facility government 3 = Clinic/ family planning 4 = Mobile Clinic *Private Sector* 5 = Private hospital/ Clinic 6 = Pharmacy 7 = Private medical doctor 8 = Mobile clinic 9 = Traditional healer 10 = Other (Specify) _____ 98 = No answer 99 = Don't know	
611	I do not want to know the result, but, the last time you were tested for HIV did you obtain the result of the test? *(State again that you do not want to know the test result)*	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	
612	Would you go for an HIV test in the future?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	*If* **Yes**, *go to* **614**

N°	QUESTIONS	ANSWERS	SKIP
613	What is the *primary* reason you don't want to go for a test? *Only one answer possible*	1 = Don't know where to go for a test 2 = Sure of not being infected 3 = Afraid of the result 4 = Afraid of the blood taking |___|___| 5 = (Afraid of) catching an infection 6 = Fear of stigmatisation 7 = Don't think testing is confidential 8 = Too expensive 9 = Other (Specify) _____ 98 = No answer 99 = Don't know	
614	Have you been given condoms by an HIV prevention program in the past 12 months?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	
615	Women only Have you been pregnant in the past 5 years?	1 = Yes 2 = No |___| 98 = No answer 99 = Don't know	*If* **No, *end interview***
616	Women only When you were pregnant did you go to an antenatal clinic?	1 = Yes 2 = No |___|___| 98 = No answer 99 = Don't know	

That is the end of the questionnaire. Thank you for taking the time to answer our questions. We appreciate your help.

Appendix 5

Coding and analysis guide

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Survey information				
Idnumber	Q01	Unique participant identification number (questionnaire number)	XXXX = number	
Hhnumber	Q02	Household number	XXXX = number	
Country	Q03	Country of survey	1 = Country	
Region	Q04	Region of survey	1 = Region1 2= Region2 Etc…	
Location	Q05	Location of interview	1= Camp 2= Surrounding area	
Site	Q06	Name of camp or community	1= Name1 2=Name2 3=Name3 Etc…	
Cluster	Q07	Cluster number	XXX = number	
Urbanrural	Q08	Urban or rural	1= Urban 2=Rural	
Interviewer	Q10	Interviewer code	XX= interviewer code	
Fieldcontrol	Q11	Field control code	XX= code	
Centralcontrol	Q12	Central control code	XX= code	
Dataclerk1	Q13	Data clerk code	XX= code	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Dataclerk2	Q14	Data clerk code	XX= code	
Starttime	Q15	Start time of interview	XX= hour XX= minutes	
Endtime	Q16	End time of interview	XX= hour XX= minutes	
Hhousehold	Q17	Head of household	1 = Yes 2 = No	
Numhhmembers	Q18	Total number of people in household	XX = number	
Numhheligible	Q19	Total number of people in household eligible for survey	XX = number	
Hhrecruit	Q20	Result of household recruitment	1 = Head agreed 2 = Head refused 3 = HH not eligible 4 = HH temporarily absent 5 = HH abandoned 6 = HH on extended travel 7 = Other	
Relationtohh	Q21	Relationship to head of household	1= Household head 2= Spouse 3= Son/daughter 4= Father/mother 5= Brother/sister 6= Other relative 7= Other person living in house who is not a relative	
Participantrecruit	Q22	Result of participant recruitment	1 =Agreed and completed 2 = Agreed and not completed 3 = Refused 4 = Absent 5 = Other	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Section 1: Background Characteristics				
A. Socio-Demographic Characteristics				
Sex	Q101	Sex of respondent/ *Total population*	1= Male 2= Female	
Age	Q102	Age of participant	XX= Age 99 = Don't know	
Birthcntry	Q103	Country of birth/ *Total population*	1=Kenya 2=Rwanda 3=Uganda 4=Tanzania 5=Congo (DRC) 6=Burundi 7=Other 98 = No answer 99 = Don't know	Responses should be revised to include only locally relevant countries.
Otherbirthcntry	Q103a	Other country of birth/ *Reported 'other' in Q103*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q103
Nationcurr	Q104	Current nationality/ *Total population*	1=Kenya 2=Rwanda 3=Uganda 4=Tanzania 5=Congo (DRC) 6=Burundi 7=Other (specify) 98 = No answer 99 = Don't know	Responses should be revised to include only locally relevant countries.
Othernationcurr	Q104a	Other current nationality/ *Reported 'other' in Q104*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q104

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Refugeecurr	Q105	Currently a refugee/ *Total population*	1=yes 2=no 98 = No answer 99 = Don't know	
Religion	Q106	Religion/ *Total population*	1=Catholic 2=Protestant 3=Moslem 4=Other	Responses should be revised to include other locally relevant religions.
Otherreligion	Q106a	Other religion/ *Reported 'other' in Q106*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q106
Education	Q107	Highest level of education completed/ *Total population*	0=Never attended school 1=Did not complete primary 2=Primary 3=Secondary 4=College 5=University 98 = No answer 99 = Don't know	
Swahili	Q108a	Ability to read Swahili/ *Total population*	1=Easy 2=Difficult 3 = Not at all	Responses should be revised to include only locally relevant languages.
Lingata	Q108b	Ability to read Lingata/ *Total population*	1=Easy 2=Difficult 3 = Not at all	Responses should be revised to include only locally relevant languages
Chinarwanda	Q108c	Ability to read Chinarwanda/ *Total population*	1=Easy 2=Difficult 3 = Not at all	Responses should be revised to include only locally relevant languages

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Otherlanguage	Q108d	Ability to read other language/ *Total population*	1=Easy 2=Difficult 3 = Not at all	Responses should be revised to include only locally relevant languages
Readany		Ability to read any of the 4 languages/ *Total population*	1=Yes 2= No	This new variable created to calculate the ability to read any language
Income	Q109	Has activity that generates income/ *Total population*	1=Yes 2= No 98 = No answer 99 = Don't know	
Incomesector	Q110	Type of income generating activity/ *Had an income (Q109=1)*	1=Agriculture 2= Trading 3=Pastoralism 4=Transport 5=Fishing 6=Crafts 7=Private services 8=Public services 9=Humanitarian or development group 10=Other 98 = No answer 99 = Don't know	
Otherincome	Q110a	Other income sector/ *Reported 'other' in Q110*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q110
Residence and mobility				
Timeres	Q111	Length of time living in current community/ *Total population*	1=Always 2= <6 months 3=6-12 months 4= 1-2 years 5 = 3-5 years 6 = More than 5 years 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Leavecntryref	Q112	Length of time since left birth country/ *Refugees only*	XX=Number of years 99 = Don't know	
Transitcntryref	Q113	Number of countries transited through since leaving home/ *Refugees only*	XX=Number of countries 99 = Don't know	
Away1mth	Q114	Been away from current home for >=1 month in past 12 months/ *Total population*	1=yes 2=no 98 = No answer 99 = Don't know	
Awayreason	Q115	Reason away >=1 mth / *Away (Q114=1)*	1= Employment 2=Trade 3=Shopping/market 4= Health care 5= School 6= Entertainment 7= Food 8= Visit relative/friend 9= Collect firewood 10= Attend religious service 11=Other 98 = No answer 99 = Don't know	Responses should be revised to include only locally relevant reasons.
Otheraway	Q115a	Other reason away/ *Reported 'other' in Q115*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q115
Visitcomm	Q116	Visit camp or surrounding community/ *Total population*	0=never 1= < once a month 2= once a month 3= Many times a month 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Visitwhy	Q117	Why last visit to other community / *Visited other community (Q116=1)*	1= Employment 2=Trade 3=Shopping/market 4= Health care 5= School 6= Entertainment 7= Food 8= Visit relative/friend 9= Collect firewood 10= Attend religious service 11=Other 98 = No answer 99 = Don't know	Responses should be revised to include only locally relevant reasons.
Othervisit	Q117a	Other reason visited/ *Reported 'other' in Q117*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q117
Marital status				
Marriedever	Q118	Ever been married/ *Total population*	1 = yes 2= no 98 = No answer 99 = Don't know	
Agefirstmarr	Q119	Age when first married / *Ever married (Q118=1)*	XX = Age 99 = Don't know	Don't know should be changed to missing to calculate mean.
Maritstat	Q120	Current relationship status/ *Total population*	1 = Currently married 2= Never married 3= Divorced/separated 4 = Widow/widower 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Monopoly	Q121	Marriage is monogamous or polygamous / *Currently married (Q120=1)*	1= Monogamous 2=Polygamous 98 = No answer 99 = Don't know	
Longtermpart	Q122	Currently living with long term partner/ *Total population*	1 = yes 2= no 98 = No answer 99 = Don't know	
B: Alcohol and drug				
Alcohol4wks	Q123	Frequency drinking alcohol in past 4 weeks/ *Total population*	1= Everyday 2= At least once a week 3= At least once a month 4= Never 98 = No answer 99 = Don't know	
Drugs12m	Q124	Took non-medically prescribed drugs past 12 months/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	
Marijuana	Q1251	Used marijuana/ *Used drugs (Q124=1)*	1= yes 2= no	1.The analysis is for a sub-population who used drugs to calculate the most commonly used drugs. Clearly show the denominator when presenting the information. 2. Responses should be revised to include locally relevant drugs.
Khat	Q1252	Used knat/ *Used drugs (Q124=1)*	1= yes 2= no	
Heroin	Q1253	Used heroin/ *Used drugs (Q124=1)*	1= yes 2= no	
Opium	Q1254	Used opium/ *Used drugs (Q124=1)*	1= yes 2= no	
Amphetamines	Q1255	Used amphetamines/ *Used drugs (Q124=1)*	1= yes 2= no	
Otherdrug	Q1256	Used other drug/ *Used drugs (Q124=1)*	1= yes 2= no	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Whatotherdrug	Q125a	What other drug used/ *Reported other drug (Q125b)*	1=Other1 2=Other2 Etc…	When feasible, each 'other drug' should be made into a separate variable.
Inject12m	Q126	Injected non-medically prescribed drugs past 12 months/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	Missing should be recoded to 'no'.
Sharedneedle12m	Q127	Used previously used needle or syringe to inject/ *Injected (Q126=1)*	1= yes 2= no 98 = No answer 99 = Don't know	Clearly show denominator when presenting the information.
C. Circumcision				
Circumcised	Q128	Circumcised/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	
Agecircum	Q129	Age circumcised / *Circumcised (Q128=1)*	XX= Age 99 = Don't know	Don't know should be changed to missing to calculate mean.
Prefpartcircum	Q130	Would prefer partner to be circumcised/ *Total population*	1= Yes, circumcised 2= No, not circumsized 3= Don't know/no preference	
Interestcircum	Q131	Would be interested in getting circumcised if afford/ safe / *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	
D. Military activity				
Militaryever	Q132	Ever involved in official or unofficial military activities/ *Total population*	1 = yes 2 = no 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Militarytime	Q133	Length of time in military activities/ *Ever in military (Q132=1)*	1 = < 6 months 2 = 6-12 months 3 = 1-2 years 4 = 3-4 years 5 = 5 or more years 98 = No answer 99 = Don't know	
Militarycurrent	Q134	Currently in military/ *Ever in military (Q132=1)*	1 = yes 2 = no 98 = No answer 99 = Don't know	
Whenleftmilitary	Q135	When left military/ *Left military (Q134=2)*	XX= Years 00= Less than one year 98 = No answer 99 = Don't know	Don't know should be changed to missing to calculate mean.
Section II: Sexually History and Risk Behavior				
A. Sexual activity				
Eversex	Q201	Ever had sexual intercourse/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	
Agefirstsex	Q202	Age at first sex/ *Ever had sex (Q201=1)*	XX= Age 99 = Don't know	Don't know should be changed to missing to calculate mean.
Condomlastsex	Q203	Used a condom at last sex/ *Ever had sex (Q201=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
B. Regular Sex Partners				
Regpart12m	Q204	Regular partner past 12 months/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	Missing should be recoded as 'no'.

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Numregpart12m	Q205	Number regular partners past 12 months/ *Had regular partner* (Q204=1)	XX=Number 98= No answer 99= Don't know	Don't know and no answer should be changed to missing to calculate mean.
Natregpart	Q206	Nationality of most recent regular partner/ *Had regular partner* (Q204=1)	1= Kenyan 2= Rwandan 3= Ugandan 4= Tanzanian 5= Congolese (DRC) 6= Burundian 7 = Other 98 = No answer 99 = Don't know	Responses should be revised to include locally relevant nationalities.
Othernatregpart	Q206a	What other nationality/ *Reported other nationality* (Q206)	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q205
Ageregpart	Q207	Age of last regular partner/ *Had regular partner* (Q204=1)	XX = Age 99 = don't know	Don't know should be changed to missing to calculate mean.
Condlastregpart	Q208	Used a condom during last sex with regular partner/ *Had regular partner* (Q204=1)	1 = yes 2 = no 98 = No answer 99 = Don't know	
C. Casual Sex Partners				
Caspart12m	Q209	Casual partner past 12 months/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	Missing should be recoded as 'no'.
Numcaspart12m	Q210	Number casual partners past 12 months/ *Had casual partner* (Q209=1)	XX=Number 98= No answer 99= Don't know	Don't know and no answer should be changed to missing to calculate mean.

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Natcaspart	Q211	Nationality of most recent casual partner/ *Had casual partner (Q209=1)*	1= Kenyan 2= Rwandan 3= Ugandan 4= Tanzanian 5= Congolese (DRC) 6= Burundian 7 = Other 98 = No answer 99 = Don't know	Responses should be revised to include locally relevant nationalities.
Othernatcaspart	Q211a	What other nationality/ *Reported other nationality (Q211)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q210
Agecaspart	Q212	Age of last casual partner/ *Had casual partner (Q209=1)*	XX = Age 99 = don't know	Don't know should be changed to missing to calculate mean.
Marstatcaspart	Q213	Marital status of last casual partner/ *Had casual partner (Q209=1)*	1=Married 2= Single 3=Divorced 4= Widow/widower 5=Other 98 = No answer 99 = Don't know	
Othermaritcaspart	Q213a	What other marital status/ *Reported other marital status (Q213)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q212

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Profcaspart	Q214	Profession last casual partner/ *Had casual partner (Q209=1)*	1= Businessperson 2= Trader 3= Student 4= Driver/truck driver 5= Housemaid 6= Pastoralist 7= Farmer 8= Military/paramilitary/ police 9= Commercial sex worker 10= Humanitarian or development worker 11= Unemployed 12 = Other 98 = No answer 99 = Don't know	
Otherprofcaspart	Q214a	What other marital status/ *Reported other profession (Q214)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q213
Alcoholcaspart	Q215	Under influence of alcohol with last casual partner/ *Had casual partner (Q209=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
Condlastcaspart	Q216	Used a condom during last sex with casual partner/ *Had casual partner (Q209=1)*	1 = yes 2 = no 98 = No answer 99 = Don't know	
Suggestcondcaspart	Q217	Who suggested condom use during sex with last casual partner/ *Used condom with last casual partner (Q216=1)*	1= My partner 2= Myself 3= Joint decision 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Whynocondcaspart	Q218	Main reason did not use condom during sex with last casual partner/ *Did not use condom with last casual partner* (Q216=2)	1= No condoms available 2= Free condoms not available 3= Too expensive 4= Partner objected 5= Don't like them 6= Used other contraceptive 7= Trust partner 8= Didn't think of using one 9= Don't know what condom is 10= Want to have child 11= Religious reasons 12= Unplanned sex 13= Didn't think was necessary 14= Other 98 = No answer 99 = Don't know	
Othernocondcaspart	Q218a	What other reason did not use condom/ *Reported other reason* (Q218)	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q217
Conscondcaspart12m	Q219	Frequency of condom use with all casual partners in past 12 months/ *Had casual partner* (Q209=1)	1= Every time 2= Frequently (more than 50% of time) 3= Sometimes (less than 50% of time) 4= Never 98 = No answer 99 = Don't know	
D. Transactional Sex				
Transpartever	Q220	Ever had sex in exchange for money/gift/favor/ *Total population*	1=yes 2=no 98 = No answer 99 = Don't know	Missing should be recoded as 'no'.
Whatlasttranssex	Q221	What exchanged for last transactional sex/ *Ever transactional sex* (Q220=1)	1= Money 2= Gift 3= Favor 4= More than one thing 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Wholasttranspart	Q222	Who was last transactional sex partner/ *Ever transactional sex (Q220=1)*	1= Refugee 2= Person from local community 3= Military/paramilitary/ police 4= Humanitarian or development worker 5= UN peacekeeper 6= Other 98 = No answer 99 = Don't know	
Othertranspart	Q222a	Who was other transaction sex partner/ *Reported other partner (Q222)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q221
Transpartpredisplaceref	Q223a	Had transactional sex partner before displacement/ *Ever transactional sex (Q220=1)* *Refugees only (Q105=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
Transpartduringdisplaceref	Q223b	Had transactional sex partner during displacement/ *Ever transactional sex (Q220=1)* *Refugees only (Q105=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
Transpartpostdisplaceref	Q223c	Had transactional sex partner after displacement/ *Ever transactional sex (Q220=1)* *Refugees only (Q105=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
Transpartprerefcomm	Q224a	Had transactional sex partner before refugees arrived/ *Ever transactional sex (Q220=1)* *Nationals only (Q105=2)*	1= yes 2= no 98 = No answer 99 = Don't know	
Transpartpostrefcomm	Q224b	Had transactional sex partner after refugees arrived/ *Ever transactional sex (Q220=1)* *Nationals only (Q105=2)*	1= yes 2= no 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Transpart12m	Q225	Transactional sex partner past 12 months/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	Missing should be recoded as 'no'.
Numtranspart12m	Q226	Number transactional partners past 12 months/ *Had transactional partner past 12m (Q225=1)*	XX=Number 98 = No answer 99 = Don't know	Don't know should be changed to missing to calculate mean.
Whatlasttranssex12m	Q227	What exchanged for last transactional sex in past 12 months/ *Had transactional partner past 12m (Q225=1)*	1= Money 2= Gift 3= Favor 4= More than one thing 98 = No answer 99 = Don't know	
Wholasttranspart12m	Q228	Who was last transactional sex partner in past 12 months/ *Had transactional partner past 12m (Q225=1)*	1= Refugee 2= Person from local community 3= Military/paramilitary/ police 4= Humanitarian or development worker 5= UN peacekeeper 6= Other 98 = No answer 99 = Don't know	
Othertranspart12m	Q228a	Who was other transaction sex partner in past 12 m/ *Reported other partner (Q228)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q227
Agetranspart	Q229	Age of last transactional partner/ *Had transactional partner past 12m (Q225=1)*	XX = Age 98 = No answer 99 = Don't know	Don't know should be changed to missing to calculate mean.

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Alcoholtranspart	Q230	Under influence of alcohol with last transactional partner/ *Had transactional partner past 12m (Q225=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
Condlasttranspart	Q231	Used a condom during last sex with transactional partner/ *Had transactional partner past 12m (Q225=1)*	1 = yes 2 = no 98 = No answer 99 = Don't know	
Suggestcondtranspart	Q232	Who suggested condom use during sex with last transactional partner/ *Used condom with last transactional partner (Q231=1)*	1= My partner 2= Myself 3= Joint decision 98 = No answer 99 = Don't know	
Whynocondtranspart	Q233	Main reason did not use condom during sex with last transactional partner/ *Did not use condom with last transactional partner (Q231=2)*	1= No condoms available 2= Free condoms not available 3= Too expensive 4= Partner objected 5= Don't like them 6= Used other contraceptive 7= Trust partner 8= Didn't think of using one 9= Don't know what condom is 10= Want to have child 11= Religious reasons 12= Unplanned sex 13= Didn't think was necessary 14= Other 98 = No answer 99 = Don't know	
Othernocondtranspart	Q233a	What other reason did not use condom/ *Reported other reason (Q233)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q232

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Conscondtranspart12m	Q234	Frequency of condom use with all transactional partners in past 12 months/ *Had transactional partner past 12m (Q225=1)*	1= Every time 2= Frequently (more than 50% of time) 3= Sometimes (less than 50% of time) 4= Never 98 = No answer 99 = Don't know	
E. Forced Sex				
Forcedsexever	Q235	Ever forced to have sex/ *Total population*	1 = yes 2 = no 98 = No answer 99 = Don't know	
Forcedsexpredisplaceref	Q236a	Forced to have sex before displacement/ *Ever forced sex (Q235=1)* *Refugees only (Q105=1)*	1= yes 2= no	
Forcedsexduringdisplaceref	Q236b	Forced to have sex during displacement/ *Ever forced sex (Q235=1)* *Refugees only (Q105=1)*	1= yes 2= no	
Forcedsexpostdisplaceref	Q236c	Forced to have sex after displacement/ *Ever forced sex (Q235=1)* *Refugees only (Q105=1)*	1= yes 2= no	
Forcedsexprerefcomm	Q237a	Forced to have sex before refugees arrived/ *Ever forced sex (Q235=1)* *Nationals only (Q105=2)*	1= yes 2= no	
Forcedsexpostrefcomm	Q237b	Forced to have sex after refugees arrived/ *Ever forced sex (Q235=1)* *Nationals only (Q105=2)*	1= yes 2= no	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Whoforcedsex	Q238	Who forced sex/ *Ever forced sex (Q235=1)*	1= Regular partner 2= Other family member 3= Non-family member 98 = No answer 99 = Don't know	
Refforcedsex	Q2391	Refugee forced sex/ *Forced sex by non-family member (Q238=3)*	1= yes 2= no	
Natforcedsex	Q2392	Person from local community forced sex/ *Forced sex by non-family member (Q238=3)*	1= yes 2= no	
Militforcedsex	Q2393	Military forced sex/ *Forced sex by non-family member (Q238=3)*	1= yes 2= no	
Humanitforcedsex	Q2394	Humanitarian or aid worker forced sex/ *Forced sex by non-family member (Q238=3)*	1= yes 2= no	
UNpkforcedsex	Q2395	UN peacekeeper forced sex/ *Forced sex by non-family member (Q238=3)*	1= yes 2= no	
Otherforcedsex	Q2396	Other forced sex/ *Forced sex by non-family member (Q238=3)*	1= yes 2= no	
Dontknowforcedsex	Q2397	Don't know forced sex/ *Forced sex by non-family member (Q238=3)*	1= yes 2= no	
Whatotherforcedsex	Q2396a	What other person forced sex/ *Reported other person (Q2396)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q2386

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Forcedsex12m	Q240	Forced to have sex past 12 months/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	Missing should be recoded as 'no'.
Numtimesforced12m	Q241	Number times forced to have sex past 12 months/ *Forced sex past 12m (Q240=1)*	XX=Number 98 = No answer 99 = Don't know	Don't know should be changed to missing to calculate mean.
Whoforcedsex12m	Q242	Who forced sex past 12m/ *Forced sex past 12m (Q241=1)*	1= Regular partner 2= Other family member 3= Non-family member 98 = No answer 99 = Don't know	
Refforcedsex12m	Q2431	Refugee forced sex 12m/ *Forced sex by non-family member (Q242=3)*	1= yes 2= no	
Natforcedsex12m	Q2432	Person from local community forced sex 12m/ *Forced sex by non-family member (Q242=3)*	1= yes 2= no	
Militforcedsex12m	Q2433	Military forced sex 12m/ *Forced sex by non-family member (Q242=3)*	1= yes 2= no	
Humanitforcedsex12m	Q2434	Humanitarian or aid worker forced sex 12m/ *Forced sex by non-family member (Q242=3)*	1= yes 2= no	
UNpkforcedsex12m	Q2435	UN peacekeeper forced sex 12m/ *Forced sex by non-family member (Q242=3)*	1= yes 2= no	
Otherforcedsex	Q2436	Other forced sex 12m/ *Forced sex by non-family member (Q242=3)*	1= yes 2= no	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Dontknowforcedsex	Q2437	Don't know forced sex 12m/ *Forced sex by non-family member (Q242=3)*	1= yes 2= no	
Whatotherforcedsex	Q2436a	What other person forced sex 12m/ *Reported other person (Q2436)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q2426
Agepersforcedsex	Q244	How old person forced them to have sex/ *Forced sex past 12m (Q240=1)*	1= Older than me 2= Young than me 3= Same age as me 99 = Don't know	
F. Anal Sex				
Analsex12m	Q245	Ever had anal sex with male or female partner/ *Total population*	1 = yes 2= no 98 = No answer 99 = Don't know	
Lastanalcondwomen	Q246	Used condom last anal sex with man/ *Anal sex past 12m (Q245=1)* *Women only (Q101=2)*	1 = yes 2= no 98 = No answer 99 = Don't know	
Analsexman12mmen	Q247	Anal sex with man in past 12 months/ *Anal sex past 12m (Q245=1)* *Men only (Q101=1)*	1 = yes 2= no 98 = No answer 99 = Don't know	
Lastanalsexmancondmen	Q248	Used condom last anal sex with man/ *Anal sex past with man 12m (Q276=1)* *Men only (Q101=1)*	1 = yes 2= no 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Analsexwoman12mmen	Q249	Anal sex with woman in past 12 months/ *Anal sex past 12m (Q245=1)* *Men only (Q101=1)*	1 = yes 2= no 98 = No answer 99 = Don't know	
Lastanalsexwomancond-men	Q250	Used condom last anal sex with woman/ *Anal sex past with woman 12m (Q249=1)* *Men only (Q101=1)*	1 = yes 2= no 98 = No answer 99 = Don't know	
Section III: Male and Female Condoms				
Heardcondom	Q301	Ever heard of condoms/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	
CondprevHIV	Q3021	Think condoms are for preventing HIV/ *Ever heard of condoms (Q301=1)*	1= yes 2= no	
Condprevpreg	Q3022	Think condoms are for preventing pregnancy/ *Ever heard of condoms (Q301=1)*	1= yes 2= no	
Condfamplan	Q3023	Think condoms are for family planning/ *Ever heard of condoms (Q301=1)*	1= yes 2= no	
Condotherpurp	Q3024	Think condoms are for other purpose/ *Ever heard of condoms (Q301=1)*	1= yes 2= no	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Dontknowcondfor	Q3025	Don't know what condoms are for/ *Ever heard of condoms (Q301=1)*	1= yes 2= no	
Whattothercondpurp	Q3024a	What other purpose for condoms/ *Reported other purpose (Q3024)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q3024
Everusedcondom	Q303	Ever used a condom/ *Ever heard of condoms (Q301=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
Knowwherecondom	Q304	Know where to obtain condom/ *Ever used condom (Q303=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
Wheregetcondom	Q305	Where usually get condoms/ *Know where to obtain condom (Q304=1)*	1= Pharmacy 2= Health facility 3= At the market 4= From friends 5= At shop 6= Community health worker 7= Other 98 = No answer 99 = Don't know	
Otherplacegetcond	Q305a	What other place obtain condoms/ *Reported other place (Q305)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q305
Getcondomalways	Q306	Can get a condom every time need one/ *Know where to obtain condom (Q304=1)*	1= yes 2= no 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Whycantgetcondom	Q307	Main constraint to getting condom/ *Can't get condom every time* (Q306=2)	1= Too far away 2= Too expensive 3= Not open at convenient times 4= Not available 5= Fear being seen 6= Health worker attitude 7= Other 98 = No answer 99 = Don't know	
Othercantgetcond	Q307a	What other constraints to getting condoms/ *Reported other constrain* (Q307)	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q307
Heardfemcond	Q308	Ever heard of the female condom/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	Missing should be recoded as 'no'.
Usedfemcond	Q309	Ever used female condom/ *Heard female condom* (Q308=1)	1= yes 2= no 98 = No answer 99 = Don't know	
Willingusefemcond	Q310	Willing to use female condom with partner / *Heard female condom* (Q308=1)	1= yes 2= no 98 = No answer 99 = Don't know	
Getfemcondom	Q311	Know where to obtain female condom/ *Heard female condom* (Q308=1)	1= yes 2= no 98 = No answer 99 = Don't know	
Section IV: Sexually Transmitted Infections				
HeardSTI	Q401	Ever heard of STIs/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Gendischarge12m	Q402	Had unusual genital discharge past 12m/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	
Ulcersore12m	Q403	Had genital ulcer or sore past 12m/ *Total population*	1= yes 2= no 98 = No answer 99 = Don't know	
SeekSTItreatment	Q404	Sought treatment when had STI/ *Had genital discharge (Q402=1) OR ulcer/sore (Q403=1)*	1= yes 2= no 98 = No answer 99 = Don't know	
Firstplacetreat	Q405	First place treated for STI/ *Sought treatment (Q404=1)*	1= Public health center 2= Private health center 3= Traditional practitioner 4= Pharmacy 5= Friend or relative 6= Other 98 = No answer 99 = Don't know	Responses should be revised to include locally relevant locations.
Otherplacetreat	Q405a	Other places sought STI treatment/ *Reported other place (Q405)*	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q405
Informpartner	Q406	Informed partner when had last STI/ *Had genital discharge (Q402=1) OR ulcer/sore (Q403=1)*	1= Yes, all of them 2= Some of them, not all 3= No, none of them 98 = No answer 99 = Don't know	
Section V: HIV Knowledge, Opinions, and Attitudes				
HeardHIVever	Q501	Ever heard of HIV/AIDS/	1 = yes 2= no 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
WhereHIVref	Q502	Think most HIV in what community/	1= My (refugee) community 2= Surrounding local community 98 = No answer 99 = Don't know	
WhereHIVnat	Q503	Think most HIV in what community/	1= My (surrounding local) community 2= Refugee community 98 = No answer 99 = Don't know	
Faithful	Q504	Can protect self from HIV by being faithful/	1= yes 2= no 98 = No answer 99 = Don't know	
Condom	Q505	Can protect self from HIV by consistently using condoms/	1= yes 2= no 98 = No answer 99 = Don't know	
Abstain	Q506	Can protect self from HIV by abstaining/	1= yes 2= no 98 = No answer 99 = Don't know	
Mosquito	Q507	Think mosquitoes transmit HIV/	1= yes 2= no 98 = No answer 99 = Don't know	Note: The correct answer to this question is 'no'.
Toothbrush	Q508	Think can get HIV from toothbrush of infected person/	1= yes 2= no 98 = No answer 99 = Don't know	Note: The correct answer to this question is 'no'.
Analsexnocond	Q509	Can get HIV from having anal sex without a condom/	1= yes 2= no 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Needle	Q510	Can protect self from HIV by not sharing needles/	1= yes 2= no 98 = No answer 99 = Don't know	
Sharefood	Q511	Think can get HIV from sharing food with infected person/	1= yes 2= no 98 = No answer 99 = Don't know	Note: The correct answer to this question is 'no'.
HealthypersHIV	Q512	Think healthy looking person can have HIV/	1= yes 2= no 98 = No answer 99 = Don't know	
Pregtransmit	Q513	Know HIV can be transmitted through pregnancy/ delivery/	1= yes 2= no 98 = No answer 99 = Don't know	
Breastfeedtransmit	Q514	Know HIV can be transmitted through breastfeeding/	1= yes 2= no 99= don't know	
Familysecret	Q515	Would want family member with HIV to be a secret/	1= yes (keep it secret) 2= no 98 = No answer 99 = Don't know	
Careillrelative	Q516	Would care for family member sick with virus /	1= yes 2= no 98 = No answer 99 = Don't know	
Contteaching	Q517	Think teacher infected with HIV should continue in job/	1= yes 2= no 98 = No answer 99 = Don't know	
Buyvegetables	Q518	Would buy vegetables from shopkeeper with HIV /	1= yes 2= no 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Adolscondom	Q519	Think adolescents should be taught about condom use /	1= yes 2= no 98 = No answer 99 = Don't know	
Personalrisk	Q520	Think are at personal risk of acquiring HIV /	1= yes, high risk 2= yes, moderate risk 3= no risk 98 = No answer 99 = Don't know	
Section VI: Exposure and Access to Interventions				
HIVinfo12m	Q601	Received information on HIV in past 12m/	1= yes 2= no 98 = No answer 99 = Don't know	
Radio	Q6021	Received information from radio/	1= yes 2= no	Missing values should be recoded as 'no'
Televisionvideo	Q6022	Received information from television or video/	1= yes 2= no	Missing values should be recoded as 'no'
Newspaper	Q6023	Received information from newspaper/	1= yes 2= no	Missing values should be recoded as 'no'
Posterpamphlet	Q6024	Received information from poster or pamphlet/	1= yes 2= no	Missing values should be recoded as 'no'
Healthfacility	Q6025	Received information from health facility/	1= yes 2= no	Missing values should be recoded as 'no'
VCTcenter	Q6026	Received information from VCT center/	1= yes 2= no	Missing values should be recoded as 'no'

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
ANCMCTCcenter	Q6027	Received information from ANC or MCTC/	1= yes 2= no	Missing values should be recoded as 'no'
Communityhlthwk	Q6028	Received information from community health worker/	1= yes 2= no	Missing values should be recoded as 'no'
Friend	Q6029	Received information from friend/	1= yes 2= no	Missing values should be recoded as 'no'
Familymember	Q60210	Received information from family member/	1= yes 2= no	Missing values should be recoded as 'no'
PLWHA	Q60211	Received information from PLHIV/	1= yes 2= no	Missing values should be recoded as 'no'
Peeroutreachwk	Q60212	Received information from peer outreach worker/	1= yes 2= no	Missing values should be recoded as 'no'
School	Q60213	Received information from school/	1= yes 2= no	Missing values should be recoded as 'no'
Placeworship	Q60214	Received information from place of worship/	1= yes 2= no	Missing values should be recoded as 'no'
Publicmtg	Q60215	Received information from public meeting/	1= yes 2= no	Missing values should be recoded as 'no'
Othersource	Q60216	Received information from other source/	1= yes 2= no	Missing values should be recoded as 'no'

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Whatothersource	Q60216a	Other source of HIV information/	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be made into separate variables.
Prefradio	Q6031	Prefer getting information from radio/	1= yes 2= no	Missing values should be recoded as 'no'
Preftelevisionvideo	Q6032	Prefer getting information from television or video/	1= yes 2= no	Missing values should be recoded as 'no'
Prefnewspaper	Q6033	Prefer getting information from newspaper/	1= yes 2= no	Missing values should be recoded as 'no'
Prefposterpamphlet	Q6034	Prefer getting information from poster or pamphlet/	1= yes 2= no	Missing values should be recoded as 'no'
Prefhealthfacility	Q6035	Prefer getting information from health facility/	1= yes 2= no	Missing values should be recoded as 'no'
PrefVCTcenter	Q6036	Prefer getting information from VCT center/	1= yes 2= no	Missing values should be recoded as 'no'
PrefANCMCTCcenter	Q6037	Prefer getting information from ANC or MCTC/	1= yes 2= no	Missing values should be recoded as 'no'
Prefcommunityhlthwk	Q6038	Prefer getting information from community health worker/	1= yes 2= no	Missing values should be recoded as 'no'
Preffriend	Q6039	Prefer getting information from friend/	1= yes 2= no	Missing values should be recoded as 'no'

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Preferfamilymember	Q60310	Prefer getting information from family member/	1= yes 2= no	Missing values should be recoded as 'no'
PrefPLWHA	Q60311	Prefer getting information from PLHIV/	1= yes 2= no	Missing values should be recoded as 'no'
Prefpeeroutreachwk	Q60312	Prefer getting information from peer outreach worker/	1= yes 2= no	Missing values should be recoded as 'no'
Prefschool	Q60313	Prefer getting information from school/	1= yes 2= no	Missing values should be recoded as 'no'
Prefplaceworship	Q60314	Prefer getting information from place of worship/	1= yes 2= no	Missing values should be recoded as 'no'
Prefpublicmtg	Q60315	Prefer getting information from public meeting/	1= yes 2= no	Missing values should be recoded as 'no'
Prefothersource	Q60316	Prefer getting information from other source/	1= yes 2= no	Missing values should be recoded as 'no'
Whatothersource	Q60316a	Other preferred source of HIV information/	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be made into separate variables.
KnowHIVtestsite	Q604	Know place person can be tested for HIV/	1= yes 2= no 98 = No answer 99 = Don't know	Missing values should be recoded as 'no'
WhereHIVtestsite	Q605	Where a person can be tested for HIV/	1= In local community 2= In refugee camp 3= In both sites 98 = No answer 99 = Don't know	

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
TestHIV	Q606	Ever tested for HIV/	1= yes 2= no 98 = No answer 99 = Don't know	Missing values should be recoded as 'no'
Whentested	Q607	When last tested for HIV/	1= In the past 12 months 2= 1-2 years ago 3= 3 or more years ago 98 = No answer 99 = Don't know	
Volmandtest	Q608	Was last test voluntary or mandatory/	1= Voluntary 2= Mandatory 98 = No answer 99 = Don't know	
Counseltest	Q609	Received counseling last HIV test/	1= yes 2= no 98 = No answer 99 = Don't know	
Wheretested	Q610	Where was site of last HIV test/	1= Hospital 2= Health facility (government) 3= Clinic/family planning center 4= Mobile clinic (public) 5= Private hospital/clinic 6= Pharmacy 7= Private medical doctor 8= Mobile clinic (private) 9= Traditional healer 10= Other 98 = No answer 99 = Don't know	
Whereothertest	Q610a	Where other site tested/	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q610

Variable recode (optional)	Variable number	Description (denominator for analysis)	Code	Notes
Getresult	Q611	Obtained result of last HIV test/	1= yes 2= no 98 = No answer 99 = Don't know	
FutureHIVtest	Q612	Would get HIV test in future/	1= yes 2= no 98 = No answer 99 = Don't know	
ReasonnoHIVtest	Q613	Reason why does not want HIV test/	1= Don't know where to go 2= Sure not infected 3= Afraid of result 4= Afraid of blood taking 5= Afraid of catching infection 6= Fear stigmatization 7= Don't think confidential 8= Too expensive 9= Other 98 = No answer 99 = Don't know	
Otherreasonnotest	Q613a	Other reason not to be tested/	1=Other1 2=Other2 Etc…	When feasible, the responses for this question should be combined with Q613
CondomHIVprev	Q614	Received a condom for HIV prevention in the past 12 months/	1= yes 2= no 98 = No answer 99 = Don't know	Missing values should be recoded as 'no'
Pregnant5yrs	Q615	Pregnant in past 5 years/	1= yes 2= no 98 = No answer 99 = Don't know	
ANCpregnant	Q616	Went to ANC clinic when pregnant/	1= yes 2= no 98 = No answer 99 = Don't know	

Conducting HIV Behavioral Surveillance Surveys among Displaced Populations and their Surrounding Communities

UNHCR
The UN Refugee Agency

Great Lakes Initiative on AIDS (GLIA)

THE WORLD BANK