Behavioural Surveillance

Surveillance of
HIV Risk Behaviours

Participant Manual

October 2009

Acknowledgments

This manual was prepared by the United States Department of Health and Human Services, Centers for Disease Control and Prevention (HHS-CDC), Global AIDS Program (GAP) Surveillance Team in collaboration with:

- The World Health Organization (WHO), Department of HIV/AIDS, Geneva, Switzerland
- the World Health Organization (WHO), Regional Office of the Eastern Mediterranean (EMRO), Division of Communicable Diseases, AIDS and Sexually Transmitted Diseases (ASD) Unit, Cairo, Egypt
- the World Health Organization (WHO), Regional Office of Africa (AFRO)
- the World Health Organization (WHO), Regional Office of South-East Asia (SEARO)
- The University of California at San Francisco (UCSF), Institute for Global Health, AIDS Research Institute through the University Technical Assistance Program (UTAP) with CDC/GAP.

Additional assistance was provided by Tulane University, School of Public Health and Tropical Medicine, New Orleans, USA, through the UTAP with CDC-GAP.

This participant manual is jointly published by HHS-CDC and UCSF.

This manual was funded by the Presidents Emergency Plan for AIDS Relief (PEPFAR) and supported by UNAIDS and the Office of the Global AIDS Coordinator (OGAC) interagency Surveillance and Survey Technical Working Group that consists of:

- United States Census Bureau
- United States Agency for International Development (USAID)
- United States Department of Defense
- United States State Department

Table of Contents

Table of Contents, continued

Introduction

How to Study This Module

What you should know before the course

This course is meant for those involved in the planning and implementation of behavioural surveillance. As a participant, you should understand the basic epidemiology of HIV/AIDS and public health surveillance.

Module structure

The module is divided into units. The units are convenient blocks of material for a single study session.

This module can also be used for self-study.

Because you already know quite a bit about HIV/AIDS, we begin each unit with some warm-up questions. Some of the answers you may know. For other questions, your answer may just be a guess. Answer the questions as best you can.

You may want to keep the warm-up exercises as a future reference. No one will see your answers but you. We will study and discuss the unit, and then you will have time to go back and change your warm-up answers. At the end of the unit, the class will discuss the warm-up questions. You can then check your work.

As you study this module, you may come across italicised terms that are unfamiliar. In Appendix B, you will find a glossary that defines these words. The glossary also contains acronyms that you may not recognise.

Module summary

The HIV/AIDS epidemic continues to grow worldwide and have a devastating impact on individuals, communities and entire countries. Behavioural surveillance measures trends in the behaviours that can lead to HIV infection. It has been shown to make important and useful contributions to facilitating how countries respond to HIV. Conducting behavioural surveillance requires co-ordination among many partners and multiple skills. Although there are many useful reference materials available for behavioural surveillance, there has not yet been a comprehensive effort to train or improve the capacity of in-country surveillance teams. This module aims to help reduce this training gap.

Appendices

More information is provided:

Appendix A, References and Further Reading Material*
Appendix B, Glossary and Acronyms*
Appendix C, Useful Links*
Appendix D, Answers to Warm-Up Questions and Case Studies

*Behavioural surveillance appendices A, B and C are used for Modules 5 and 6 of this curriculum.

Additions, Corrections, Suggestions

Do you have changes to this module? Is there additional information you would like to see? Please write or email us. We will collect your letters and emails and consider your comments in the next update to this module.

Mailing address:

Global Health Sciences
Prevention and Public Health Group
University of California San Francisco
50 Beale Street, Suite 1200
San Francisco, California 94105 USA

modules@psg.ucsf.edu

Unit 1

Behavioural Surveillance

Overview

What this unit is about

In this unit, you will learn basic information about surveillance, its purpose, history and key considerations.

Warm-up questions

1. True or false? One-time cross-sectional surveys can be considered surveillance.

   True False

2. _____ surveillance involves regular, repeated cross-sectional surveys collecting data that can be compared over time to HIV risk behaviours and other relevant issues.

   a. Behavioural
   b. Biological

3. Which of the following can be a use of behavioural surveillance?

   a. To explain changes in HIV prevalence over time
   b. To provide information for prevention programmes
   c. To raise the awareness of HIV among policy-makers
   d. To provide an early warning for areas and groups at risk for transmission of infection
   e. All of the above

4. True or false? Surveillance is a useful tool for evaluating specific HIV/AIDS interventions.

   True False

5. True or false? In a generalised epidemic everyone is at equal risk of infection.

   True False

Warm-up questions, continued

6. _____ *sites* are facilities such as STD clinics, antenatal care clinics, blood donation centres, drug treatment programmes, prisons and needle exchange programs.

   a. Sentinel
   b. Community

7. Which of the following is the definition of linking behavioural and biological data?

   a. Collecting HIV, STI and behavioural data from the *same individuals* at the same time
   b. Collecting HIV, STI and behavioural data from the *same source population* at different times
   c. Analysing HIV, STI and behavioural data from *similar source population*, using whatever data are available
   d. Reporting behavioural and biological surveillance together
   e. All of the above

8. Collecting _____ level data provides more detailed information but requires larger sample sizes and thus more time and money.

   a. National
   b. Sub-national

Introduction

What you will learn

By the end of this unit, you should be able to:

- define surveillance
- identify the uses of behavioural surveillance
- identify issues to consider when designing a surveillance system
- identify the steps required to achieve a sustainable surveillance system.

Surveillance is the systematic, regular and ongoing collection and use of data for public health action. Although they are often the beginning of a surveillance system, one-time cross sectional surveys are not considered as surveillance. HIV/AIDS surveillance is divided into biological and behavioural surveillance.

Behavioural surveillance involves regular, repeated cross-sectional surveys that collect data on HIV risk behaviours and other relevant issues. These surveys can be compared over time.

In *biological surveillance*, biological samples are collected and tested for HIV and other related illnesses such as STIs and TB.

Cross-sectional surveys collect information from a selected sample of a target population at one point in time or over a short period of time. In surveillance, the same survey or a similar survey is repeated with the same target population (but a different sample of people) at regular intervals. This enables us to explore behavioural changes over time.

Uses of behavioural surveillance

Uses of behavioural surveillance include the following:

- To provide an early warning of groups and areas that infection is likely to spread in and between.
- To explain changes in HIV prevalence over time. Without behavioural data, biological surveillance data are difficult to interpret. For example:

- Does a stable or falling HIV prevalence mean:
    - there are fewer new infections?
    - more deaths?
    - the population being tested has changed over time?

Uses of behavioural surveillance, continued

- Does a rising HIV prevalence mean:
  - prevention programmes are failing?
  - life expectancy is increasing because of treatment programmes?
  - the epidemic hasn't reached the stage when people are dying?
  - the testing population has changed over time?

- To provide information for developing prevention programmes. Measuring the prevalence of HIV alone does not provide all the information needed to design effective policy and programmes. Behavioural data allows us to identify the populations and behaviours that are driving the epidemic and that should be targeted in programmes.

  Note: surveillance does not provide much information about how to target the groups and behaviours, that requires objective research.

- To monitor and evaluate the impact of prevention programmes. Surveillance can be a useful tool for monitoring and evaluating HIV/AIDS prevention programmes that target the populations or the geographical areas included in surveillance. The national monitoring and evaluation strategies for HIV/AIDS should, therefore, incorporate indicators derived from behavioural surveillance data. However:

  - Without adaptation, surveillance only provides evidence for the impact of HIV programmes as a whole and not for the impact of specific interventions or specific programme elements. Although surveillance can be adapted to evaluate specific interventions (by adding questions relating to exposure to the specific interventions) this must be done with care. Appropriate adaption ensures the focus of surveillance is not deflected away from detecting and measuring risk behaviours.

  - While surveillance is useful for evaluating programmes, like most evaluation methods, it does not provide conclusive evidence that the programme *caused* any observed changes in behaviour. Any observed change may have occurred without the programme due to other factors.

- To reinforce the findings of biological surveillance.

- To raise the awareness of HIV among policy-makers.

Designing a Behavioural Surveillance System

Considerations

When designing a behavioural surveillance system, you should consider:

- whom to include in surveillance
- where to access the surveillance populations
- how to link biological and behavioural surveillance data
- how to ensure that surveillance is appropriate for the context

Whom to include

The current guidelines about whom to include in surveillance differ according to the severity of the epidemic in the affected country. Epidemics can be broadly classified into low grade, concentrated and generalised epidemics.

- *Low-level* - Prevalence of HIV is consistently below 5% in any "high-risk groups" *and* below 1% in the "general population."

- *Concentrated* - Prevalence of HIV has surpassed 5% on a consistent basis in one or more "high-risk groups," but remains below 1% in the "general population."

- *Generalised* - Prevalence of HIV has surpassed 1% in the "general population."

The general guidelines for whom to include in surveillance for each epidemic state are shown in Table 1.1.

Table 1.1. Whom to include in surveillance by epidemic state.

State of the epidemic	Biological surveillance (annually if feasible)	Behavioural surveillance
Low-level	High-risk groups	High-risk groups annually, general population every 3-5 years
Concentrated	High-risk groups, general population	High-risk groups annually, general population every 3-5 years
Generalised	High-risk groups, general population	High-risk groups annually, general population annually

Discussing the table

Looking at Table 1.1, answer the following question:

What state of the epidemic is your country in?

General population and high-risk group surveillance

General population surveillance measures HIV risk behaviours in a sample of people selected to represent the people living in a region or nation. The surveillance can be restricted to certain ages (for example, young people 15-24 years old) or genders.

High-risk group surveillance measures HIV risk behaviours in groups whose behaviours, occupations or lifestyles could expose them to higher risk of acquiring and transmitting HIV than the rest of the population. These groups are often important in establishing, accelerating or sustaining the HIV epidemic.

Common high-risk groups considered for inclusion in behavioural surveillance are shown in Table 1.2.

Table 1.2. High-risk groups often considered for inclusion in behavioural surveillance.

High-risk groups often considered for inclusion in behavioural surveillance
▪ injecting drug users ▪ university students ▪ sex workers and their clients—for example, truck drivers or mine workers ▪ men who have sex with other men ▪ uniformed personnel (police, border personnel and military) ▪ migrant labourers ▪ young people

Discussing the table

Looking at Table 1.2, discuss the following question:
Do you think all of the groups listed above are high-risk groups in your setting? Are there groups missing from the table?

Why include these groups?

Here are two frequent questions asked about surveillance:

- Why should we perform general population surveillance in concentrated epidemics where the general population is at low risk?
- Why should we perform high-risk group surveillance in generalised epidemics where the whole population is at high-risk?

What are your thoughts on this?

High-risk group surveillance in a generalised epidemic: In a generalised epidemic, conduct surveillance in sub-populations where HIV is concentrated. Include their partners, because even in a generalised epidemic, not everyone in the population is at equal risk of HIV or has an equal role in the spread or maintenance of the epidemic. The behaviours of these sub-populations may continue to drive a generalised epidemic and are therefore important to monitor and to target for interventions.

General population surveillance in a concentrated epidemic: In a concentrated epidemic, general population surveillance helps you to understand the potential for HIV to spread beyond the groups where it is concentrated. Therefore, the surveillance acts as an early warning system by allowing you to explore the following:

- size of the sub-populations in which HIV is concentrated
- links (bridges) between the sub-populations and their partners and the 'general' population
- level of risk behaviours in the 'general' population

The need for general population surveillance in low-level epidemics remains controversial. Some people argue that it is not cost-effective to set up large-scale general population surveys in low-level epidemics. This should not prevent countries from conducting intermittent surveys.

Where to access groups

The populations included in surveillance can be accessed either in 'sentinel sites' or in the community.

Sentinel sites are facilities such as STD clinics, antenatal care clinics, blood donation centres, drug treatment programmes, prisons and needle exchange programmes.

Community sites are locations in the community, such as households or brothels.

Sentinel surveillance was initially used in the 1980s, when surveillance relating to HIV/AIDS was in its early stages and focused on unlinked anonymous blood testing. Sentinel surveillance is often more convenient, cheaper and has fewer ethical implications than population-based surveillance. The main drawback of sentinel surveillance is that we do not know who the people identified through the sentinel sites represent or how they change over time. All we know is that these are people who come into contact with the sites. In contrast, when community-based surveillance uses rigorous sampling techniques, the people whom the sample represents can be clearly defined.

Note: Using antenatal clinics as sentinel sites is less problematic, as women attending antenatal care are considered to be a reasonable proxy for the general population.

In *unlinked anonymous testing*, a sample of blood originally collected for other purposes is tested for HIV. The person whose blood is taken does not know that his or her blood will be tested for HIV. All information that could identify the person is removed from the sample so that the results of the test cannot be linked back to that person.

How to link biological and surveillance data

We have already discussed the importance of behavioural data for the interpretation of biological data over time. To ensure data are complementary and useful, behavioural and biological surveillance are best planned together.

When you are planning, decide how behavioural and biological data are best linked. Linking can involve several variables. The following examples are types of behavioural and biological linking:

- collecting HIV, STI and behavioural data from the same individuals at the same time
- collecting HIV, STI and behavioural data from the same source population at different times
- analysing HIV, STI and behavioural data from similar source populations, using whatever data are available
- reporting behavioural and biological surveillance together.

There are advantages and disadvantages for each type of linking. The decision on how to best link data should be country-specific. At the very least, behavioural and biological surveillance should be presented in a single report. Table 1.3 next page outlines the advantages and disadvantages of each type of linking.

Regardless of what type, linking should be done for trends over time rather than for one point in time. This is because behavioural and biological data linked at one point in time is impossible to interpret due to the temporal relationship between HIV and risk behaviours. Recent risk behaviours are not necessarily related to HIV status, as infection may have occurred some time ago. The relationship between behaviours and STIs may also be unclear.

Table 1.3. Advantages and disadvantages of different types of linking.

Type of linking	Advantages	Disadvantages
Collect data from same individuals	▪ Helps in understanding biological data over time and ensures behavioural and biological samples are alike. ▪ Behavioural and biological data can be used to reinforce each other's findings. ▪ Can be cheaper if data collection activities are combined. ▪ More convenient, especially where access is an issue. ▪ Only disrupt community once.	▪ Temporal relationship between behaviour and HIV/AIDS/STI makes individual analysis difficult. ▪ Blood draw and longer interviews may increase refusal rates. ▪ Required frequency, sample size and sampling strategies (for example, sentinel sites versus community sampling) may differ for biological and behavioural surveillance. ▪ Difficult to find fieldworkers with the skills for both behavioural and biological data collection.
Collect data from same source population	▪ Helps in understanding biological results over time, samples likely to be similar. ▪ Behavioural and biological data can be used to reinforce each other's findings. ▪ No need to compromise on fieldworker skills, frequency of surveillance, sample size or sampling strategies. ▪ Refusal rate for behavioural surveillance not increased by blood being drawn for biological surveillance.	▪ Temporal relationship between behaviour and HIV/AIDS/STI makes individual analysis difficult. ▪ No guarantee that behavioural and biological surveillance samples are alike. ▪ No savings in cost or convenience. ▪ Community disrupted more than once.
Analyse data for similar population	▪ Helps in understanding biological results over time. ▪ Behavioural and biological data can be used to reinforce each other's findings. ▪ No need to compromise on fieldworker skills, frequency of surveillance, sample size or sampling strategies. ▪ Refusal rate for behavioural surveillance not increased by blood being drawn for biological surveillance.	▪ Temporal relationship between behaviour and HIV/AIDS/STI makes individual analysis difficult. ▪ Using different source populations means that behavioural and biological surveillance samples may not be alike. ▪ No savings in cost or convenience.
Produce a joint report	▪ Provides data users with all the information they need in one document. ▪ No need to compromise on fieldworker skills, frequency of surveillance, sample size or sampling strategies. ▪ Refusal rate for behavioural surveillance not increased by blood being drawn for biological surveillance.	▪ Unlikely that the populations are similar so no integrated analysis. ▪ No savings in cost or convenience. ▪ Behavioural data adds little to understanding the findings of biological surveillance.

Discussing the table

Looking at Table 1.3, discuss the following questions:

a. What are some advantages of collecting data from the same source population?

b. What are your recommendations and justifications for linking in your country?

How to ensure surveillance is appropriate

There is no one-size-fits-all way of designing a surveillance system. The surveillance system needs to be designed to fit the specific features of your country's epidemic and health system. The considerations that we discuss in this training are not "rules." Surveillance systems require adaptation to fit the specific needs of each country.

Steps in Conducting Behavioural Surveillance

This section introduces the steps in conducting behavioural surveillance. You will learn how the different aspects of behavioural surveillance feed into each other and gain an understanding of the whole surveillance process.

The steps are listed below then explained in detail.

1. Identify a co-ordinating body.
2. Agree on the purpose of surveillance.
3. Establish criteria for selecting populations and geographic coverage areas.
4. Gather information to help with the decision about populations and geographic locations in order to guide survey implementation.
5. Finalise selection of sub-populations and geographic locations for surveillance.
6. Choose a sampling design.
7. Develop survey protocol.
8. Build a sampling frame.
9. Conduct surveillance.
10. Analyse and use data.
11. Plan for next round of surveillance.

Step 1: Identify a co-ordinating body

The purpose of the surveillance co-ordinating body is to provide guidance and serve as an over-all decision-making committee for the surveillance system. Examples of the responsibilities of a surveillance co-ordinating committee include:

- defining the purpose of surveillance
- ensuring the surveillance system is set up to meet the data needs of the country
- identifying funding sources
- advocating to policy-makers and stakeholders about the importance of surveillance
- facilitating co-ordination between surveillance partners
- making final decisions about the populations and geographical areas included in surveillance
- monitoring the progress of the surveillance process
- supporting data entry and subsequent interpretation and conclusions
- maximising data dissemination and use.

Surveillance committees may already exist, but may lack sufficient infrastructure to function appropriately. The personnel included on the committee should reflect the country's needs and experiences. Ideally, the Ministry of Health or National AIDS control programme should convene committees.

The committees should include the various national and international bodies whose interests are served by the surveillance system, such as representatives from:

- key government ministries
- multi-lateral agencies
- donor agencies
- local and/or international NGOs
- members of the target population/NGOs that represent them
- agencies selected to implement surveillance

Step 2: Agree on the purpose

The first task of the surveillance committee is to agree on the purpose of surveillance (outlined earlier in this module) and establish the data needs of that country.

While it is clear that consensus and co-ordination are important, they are often difficult to achieve, as stakeholders can have different ideas about the purpose and practicality of surveillance.

Ensuring consensus involves regular meetings and discussions among many organisations, which can be difficult to co-ordinate. The government should play a central role in this process.

Step 3: Establish criteria for selections

Before each round of surveillance, the surveillance populations and geographical areas should be reassessed. The addition of new surveillance groups may be required if new risk groups emerge or surveillance capacity increases. Old groups may need to be phased out if data is ineffectual. This should be done using the following set criteria:

- Criteria for selecting populations: The selection of surveillance populations should be based on a solid understanding of the epidemic dynamics in a country. The selected populations should be considered central to monitoring and tracking the future course of the epidemic. It is not sufficient to select surveillance populations through anecdotal evidence. Often, stakeholders want to include their priority groups (for example, in order to evaluate their own interventions) or because of political pressure to include or exclude groups (for example, a government may not want to acknowledge intravenous drug use as a problem).

- Criteria for selecting geographical coverage areas: The geographical coverage area of surveillance should be determined based on what areas of the country and what levels of data will be most meaningful in understanding the epidemic. The main criterion for selecting geographical coverage areas is whether data is required at a national or sub-national level.

  Note on national versus sub-national estimations: National level data provide estimates for the whole country. Sub-national level data provide more detailed information about different epidemics in different regions/provinces.

Step 3: Establish criteria for selections continued

When making decisions about geographical coverage areas, it is important to remember that you can carefully aggregate provincial figures to obtain a rough national figure. You can rarely disaggregate a national number to get detail at the provincial level. The downside of collecting sub-national level data is that it requires larger sample sizes and thus more time and money.

Step 4: Gather information

Once the criteria for selecting populations and geographical areas have been established, it is time to actually make the selection. Begin the process by assessing what is currently known about the national epidemic by reviewing previous research. Once this review has been completed and a list of potential sub-populations and geographical hotspots has been completed, a series of field assessments is required to find out:

- whether members of the proposed sub-populations do in fact engage in risky behaviours
- whether members of the population exist in sufficient numbers to merit inclusion in the surveillance system
- how and where members of the sub-population can be assessed and sampled in a systematic fashion
- how the sub-population can be operationally defined and what the eligibility criteria for inclusion in surveillance should be
- the general willingness of potential respondents to participate in surveillance surveys.

Step 5: Finalise selections

All the information collected in the review of previous research and in the field assessments needs to be synthesised for final population and geographic selection by the surveillance co-ordinating committee.

Step 6: Develop a sampling design

Due to the size of the surveillance populations, individual interviews are difficult. Sampling methods account for an unbiased and statistically accurate composition of the desired variables. There are several ways of sampling individuals, and the most appropriate way of ensuring that unbiased and precise estimates are obtained depends on the characteristics of the population of interest.

Step 7: Develop a survey protocol

Formulating and documenting the surveillance protocol is essential for the collection of high quality data that meets the needs of the country. Developing the protocol clarifies the aims of the surveillance system and ensures a thought out and planned protocol for all the elements of surveillance. The elements that should be included in the surveillance protocol are shown in Table 1.4 on the next page.

Use of a protocol ensures documented surveillance procedures are employed. The consistent methodology produces comparable trend data across studies, regardless of changes in personnel between data collection rounds.

Several terms and concepts are used in Table 1.4.
In element 4 of the table:

- *Operational definitions of target populations* are definitions that are useful for sampling and fieldwork purposes. For example, a definition of sex worker should clearly define what constitutes a sex worker in terms of duration of selling sex, form of payment, type of venue where they work, etc.

- An *indicator* defines an aspect of behaviour that is key to the spread of HIV. Indicators provide a way to track changes in behaviours over time and provide a way to compare levels of risk behaviours between different population groups. International indicators were developed by UNGASS to help monitor global progress in relation to HIV/AIDS: http://www.measuredhs.com/hivdata/

In element 5 of the table, data analysis and data use are sometimes only considered once data collection is completed, but planning for these at the start of the surveillance process is the only way to ensure that appropriate data is collected and that data is used in a timely and effective manner.

Step 7: Develop a survey protocol, continued

Table 1.4. The elements of a behavioural surveillance protocol.

Element	Description
1	A brief description of the HIV/AIDS situation in the country and of the existing surveillance system
2	The objectives of the surveillance system, including how surveillance is integrated into the national monitoring and evaluation strategic plan
3	A justification of the selection of sub-populations and geographic locations for surveillance and the frequency with which the surveillance will take place
4	A description of the methodologies to be used: ▪ operational definitions of the target population ▪ key indicators ▪ sampling approach for each target population ▪ sample size calculations for each target population ▪ development and validation of tools and instruments ▪ selection and training of interviewers, supervisors, data-entry clerks, etc. ▪ data collection procedures ▪ lab test kits, lab performing tests ▪ quality control
5	A data management and data analysis plan
6	The report writing and dissemination strategy and a description of how the data are expected to be used
7	A discussion of ethical considerations, including consent forms
8	The roles and responsibilities of collaborating partners
9	The timeline
10	The budget and available resources

Discussing the table

Looking at Table 1.4, answer the following questions:

a. What are some of the methodologies you can use?
b. What additional details should you include when determining the objectives of the surveillance system?

Step 8: Build sampling frame

Some sampling approaches require a list of units, or sampling frame, from which the sample is selected. If a sampling frame is required, it is usually necessary to map or enumerate the population of interest.

Step 9: Conduct surveillance

The next step is to actually carry out the surveillance protocol. If elements of the protocol are changed during implementation, this should be documented. This ensures that data collected at the next round can be collected in a comparable manner.

As the protocol is implemented, it can be useful to collect information from fieldworkers and supervisors regarding problems they encountered during data collection, as well as potential solutions, so data collection can be improved for future rounds.

Step 10: Analyse and use data

Prior to data collection, specific analyses to be run on the data should be identified. This begins with assigning personnel with appropriate data management, analysis skills, identifying software, defining indicators and making mock-ups of key tables and graphs. This planning ensures that all the required data and no unnecessary data are collected. It also ensures that data are analysed appropriately and are ready for dissemination in a timely manner.

Careful forethought into the dissemination of data is a vital component of proper analysis. Equal time should be dedicated to proper dissemination of data as planning and collection. The subsequent results of surveillance should be shared at a national, regional and district level.

Step 11: Plan for future rounds

Although the surveillance steps have been presented as linear, the process should be viewed as a cycle. The purpose of surveillance is to collect trend data, and this requires that similar methods are used for each round of data collection. Surveillance systems require adaptation and modification. As the current knowledge and intricacies of HIV/AIDS develop, and as the epidemic changes throughout a given country, proper adjustments to data collection must be made. Keep in mind the consistency of the data themes must be upheld for future analyses.

Summary

Surveillance is the systematic, regular and ongoing collection and use of data for public health action. Behavioural surveillance involves regular, repeated cross-sectional surveys collecting data that can be compared over time with HIV risk behaviours and other relevant issues. When designing a behavioural surveillance system, there are several issues you should consider. You should also follow the steps required to achieve a sustainable surveillance system.

Unit 1 Exercises

Warm-up review

Take a few minutes now to look back at your answers for the warm-up questions at the beginning of the unit. Make any changes you would like to make.

Small group discussion

Get into small groups by country, region or province to discuss the following questions.

1. How is surveillance organised/co-ordinated in your settings? What works and doesn't work in the organisational structure?

2. Look at the diagram on slide 34. This diagram was developed using behavioural surveillance data from Asia, where the majority of countries have concentrated epidemics. The diagram helps us understand what behaviours are driving the HIV epidemic, the size of high-risk groups and their links to the general population. The ovals show the different population groups and their size, while the arrows show the links between the populations and the strength of the links. Describe how the information shown in the diagram could be applied to your country.

Apply what you've learned/ case study

As an individual, answer the following questions. We will discuss the answers in class.

1. When discussing guidelines for whom to include in surveillance for each epidemic state, define what is meant by the "general population" and by "high-risk group surveillance"?

2. Name a group you consider high risk. Why do you consider them high risk? Describe the group and tell why you think it is high risk.

Notes

Unit 2

Measures and Indicators for Behavioural Surveillance

Overview

What this unit is about

In this unit, you will learn about selecting appropriate measures and indicators for behavioural surveillance.

Warm-up questions

1. What are two characteristics of a good indicator?

2. True or false? Behavioural surveillance indicators should measure aspects of behaviours that are key components in the spread of HIV. Circle your answer.

   True False

3. When should behavioural surveillance indicators be selected during behavioural surveillance? Circle your answer below.

   a. during planning
   b. during analysis

4. True or false? The two most difficult issues when creating behavioural surveillance indicators are defining the behaviours themselves and defining the time period to which the indicator should refer.

   True False

5. True or false? Behavioural surveillance indicators do not need to be consistent over time.

   True False

6. The time reference period of an indicator should be determined by its _____.

   a. prevalence
   b. frequency

Introduction

What you will learn

By the end of this unit, you should be able to:

- understand the characteristics of a good indicator
- select indicators for behavioural surveillance
- understand the main methodological obstacles when choosing indicators for behavioural surveillance
- discuss the difference between standardised and locally adapted indicators.

Good indicators

An indicator is simply a measure of something. A good indicator:

1. Measures something of relevance to the topic (the measure serves a use).
2. Measures the item of interest accurately.
3. Is easy interpreted and defined in clear terms. For example, if we were interested in measuring unsafe sex we would need to:

    - Define unsafe sex. (For example, 'sex without condom use with a commercial partner.')
    - Define the population to whom the measure can be generalised. (For example, 'men aged 15-49.')
    - Define the time period to which the measure refers. (For example, 'last sex.')

4. Can be compared across different population groups and across time. This aspect requires definitions and field methodologies are consistent for collection of the indicator.
5. Is feasible to collect in terms of effort and cost.

Selecting Indicators For Behavioural Surveillance

Behavioural surveillance indicators should measure aspects of behaviours that are key components in the spread of HIV, including:

- behaviours that determine the likelihood that an uninfected person will come into contact with an infected person (number and type of sexual partners, patterns of needle exchange, etc.).
- behaviours that determine the likelihood that transmission of HIV will occur if contact with an HIV infected person comes about (level of condom use, equipment sharing practices, etc.).

Note: The likelihood of transmission is also determined by other factors, such as the presence of other STIs.

Indicators should be selected during the planning stage of behavioural surveillance. This helps ascertain the surveillance data needs of the country. It also ensures that all the data required to construct the selected indicators is collected.

Assessing HIV risk

The selection of indicators should be determined by the specific data requirements of each country. Essential indicators for the general population, IDUs and CSWs are shown in Table 2.1. (Learn about other common indicators at http://www.measuredhs.com/hivdata/ind_tbl.cfm).

It is also common for data to be collected on factors that promote or reduce high-risk behaviours (for example, alcohol or drug use). It is also useful to collect background information so indicators can be compared across different socio–demographic profiles.

Table 2.1. Essential Indicators.

General population	IDUs	CSWs
▪ proportion who had commercial sex in past year ▪ frequency of commercial sex in past year ▪ proportion who had non-regular/casual partners in past year ▪ frequency of non-regular/ casual partners in past year ▪ last time and consistent condom use by partner type ▪ proportion who injected drugs in past year	▪ proportion who shared needles last time ▪ proportion who did not use clean needles consistently in past week (or other time reference period)	▪ last time and consistent condom use with clients ▪ proportion who injected drugs in past year

Discussing the table

Looking at Table 2.1 on the last page, discuss the following questions:

a. What are some of the essential indicators for commercial sex workers?

b. Which of these essential indicators do you currently use in your country?

Methodological issues

The two most difficult issues defining behavioural surveillance indicators are:

- defining the behaviours themselves
- defining the time period during which the indicator should refer

Defining behaviours

A good indicator is easy to interpret and is defined in clear terms. Small discrepancies in defining an indicator may later confound how to compile and analyze the subsequent data.

For example, an indicator measuring consumption of commercial sex by married men could be defined as the proportion of:

- married men who go to a sex worker
- men who go to sex workers who are married

The first indicator gives us an idea about the size of the link between commercial sex workers (CSWs) and married men. If a large proportion of married men use CSWs, then commercial sex could potentially be important factor driving the epidemic. The second indicator does not tell us anything about the role of CSWs in driving the epidemic. We do not know if the use of CSWs is high, only that among those people who use CSWs, a certain proportion are married.

Defining behaviours, continued

Definitions also need to be consistent across time so trends can be measured effectively. For example, defining consistent condom use among sex workers as 'condom use during every sex act in the last week' in the first round of surveillance and as 'condom use with every client in the last week' in the second round would result in incomparable indicators.

Similarly, it would be difficult to measure trends in intravenous drug use (IDU) behaviour if in the first surveillance round; an IDU is defined as someone who has ever injected drugs, and in the second round defined as someone who injected drugs in the last six months. The first round would include people who used drugs a long time ago and subsequently stopped. The second round refers to people who have used drugs more recently.

Reference periods

An indicator's reference (time) period can affect how accurately the indicator is measured. The reference period should be determined by the frequency of a type of behaviour but should be as short as possible to reduce the likelihood that respondents will not remember. If a type of behaviour is very common, respondents may have difficulty remembering how they behaved. A shorter reference period may be more appropriate. If the behaviour is very rare, measurement over a short reference period may not capture exposure adequately.

For example: It would be difficult for most current IDUs to remember how many times they injected in the last year. If the last year is used as a reference period, it would result in inaccurate information. If a reference period that was too short was used (for example, in the last hour) the IDU would remember accurately how many times they had injected. Most likely though, the most common response would be*' not at all'* and the data on risk behaviours would be compromised.

Standardised Versus Locally Adapted Indicators

We have already discussed the importance of using standardised indicators over time in order to obtain an accurate measure of trends. It can also be important to use standardised indicators across regions or internationally to allow comparability. The UNGASS indicators at http://www.measuredhs.com/hivdata/ind_tbl.cfm are a good example of international indicators. These are designed to help monitor global progress in relation to HIV/AIDS, but their definitions and recall periods may not be those useful or meaningful at a country level. Generally, the needs of the country should come before international needs. Indicators should be locally appropriate in terms of reference periods and definitions or they will be obsolete.

Summary

Behavioural surveillance indicators should measure aspects of behaviours that are key components in the spread of HIV. Indicators should be selected during the planning of behavioural surveillance. This increases the facilitation of the surveillance data needs of the country. It also ensures that all the data required to construct the selected indicators is collected. It can also be important to use standardised indicators across regions or internationally to allow comparability.

Unit 2 Exercises

Warm-up review

Take a few minutes now to look back at your answers for the warm-up questions at the beginning of the unit. Make any changes you want to make.

Small group discussion

Get into small groups by country, region or province to discuss these questions.

1. What things we would need to define in order to measure high-risk sex in the general population?

2. Discuss the following questions and how well the definitions work in the context of your country:

   a. How can we define *regular* and *non-regular* sexual partners?

   b. Can single people have regular partners? What if they don't live together?

   c. What is meant by *commercial sex*?

Apply what you've learned/ case study

Try this case study individually. We'll discuss the answers in class.

1. What do the following indicators measure? Which is more useful?
   a. Married men who go to sex workers

   b. Men who go to sex workers who are married.

2. Discuss the differences between substance users and illicit drug users.

Notes

Unit 3

Survey Methods

Overview

What this unit is about

In this unit, you will learn about selecting and adapting instruments and methodologies for behavioural surveillance.

Warm-up questions

1. What is measurement error?

2. In observational studies, which of the following is a source of measurement error? Circle your answer.

   a. interviewer error
   b. respondent error
   c. questionnaire faults
   d. all of the above

3. True or false? Bias can be controlled for during data analysis.

   True False

4. True or false? A face-to-face interview is the best data collection method.

   True False

5. Which of the following is an advantage of using the self-administered data collection method?

   a. inexpensive to administer
   b. no literacy requirement
   c. date entry step eliminated

Warm-up questions, continued

6. Which of the following are advantages of adapting survey questions from surveys that have already been successfully implemented?

    a. builds on current best practice of how questions can be best expressed
    b. saves time and money
    c. ensures consistency with other available data sources
    d. eliminates the need to pre-test the questionnaire
    e. all of the above

7. What is one solution to the problem of interviewer safety when working with hard-to-reach groups?

8. True or false? When having difficulty reaching members from hard-to-reach groups to show up for an interview, one solution is to use incentives.

    True False

Introduction

What you will learn

By the end of this unit, you should be able to:

- select appropriate data-collection methods
- design and adapt survey instruments
- understand measurement error and how to reduce it
- anticipate potential fieldwork difficulties and identify solutions
- select appropriate fieldworkers and supervisors

Data Collection Methods

The type of data collection method can influence how respondents answer questions and thus, influence how accurately the data collected measures the behaviour of interest. This is particularly true of sensitive data, when ensuring the privacy of the responses can be very important. Data collection methods commonly used in behavioural surveillance include the following:

Face-to-face interviews—An interviewer reads the questions to the respondent. The respondent gives an oral response, which the interviewer records on the questionnaire.

Self-administered questionnaire—The respondent is given or mailed the questionnaire, which the respondents read and complete themselves.

Computer-assisted—The respondent uses headphones to listen to questions and completes the questionnaire on a computer.

The most appropriate method will be context-specific. Each data-collection method has advantages and disadvantages. These are shown in Table 3.1 next page. There is no conclusive evidence that any one method is better, overall, than the others.

Data Collection Methods, continued

Table 3.1. Advantages and disadvantages of various data collection methods.

Method	Advantage	Disadvantage
Face-to-face	▪ no literacy requirement ▪ interviewer-respondent rapport can increase truthfulness ▪ high response rate ▪ allows flexibility in length/style of survey	▪ labour intensive ▪ low degree of privacy and anonymity ▪ interviewers may distort answers
Self-administered	▪ higher degree of privacy and anonymity ▪ inexpensive to administer	▪ respondents must be literate ▪ higher non-response rate ▪ questionnaire must be short
Computer assisted	▪ higher degree of privacy and anonymity	▪ respondents must be literate and at ease with computers ▪ requires expensive equipment
Audio assisted	▪ higher degree of privacy and anonymity ▪ data entry step eliminated	▪ requires expensive equipment

Discussing the table

Looking at Table 3.1, discuss the following questions:

a. What are some of the disadvantages of using the computer-assisted method?

b. Which method has the advantage of a high response rate?

Survey Instruments

Available instruments

You have already learned that a characteristic of a good indicator is that it measures the behaviour of interest accurately and accuracy can be improved by appropriate question phrasing. Luckily, we are not starting from scratch and can adapt our questions from surveys that have already been successfully implemented. This:

- builds on current best practice of how questions can be best expressed
- saves time and money
- ensures consistency with other available data sources.

It is still essential to pre-test and pilot questions that have been borrowed from other surveys to adapt them for each context.

There are two general population survey instruments that are widely used. They use solid sampling procedures, thorough statistical analysis, and are considered a reliable source of behavioural data:

- demographic and health surveys (MACRO)
- Multiple Indicator Cluster Survey (UNICEF).

Behavioural surveillance surveys (FHI) are also widely used surveys for high-risk groups.

Table 3.2 shows the most common HIV/AIDS relevant content of each survey and Internet links for accessing the questionnaires.

Table 3.2. Content of surveys useful for behavioural surveillance.

Survey	Content
Demographic and health surveys (DHS), AIDS Indicator Survey (AIS), MACRO International	▪ Knowledge and source of knowledge of AIDS and other STIs ▪ Knowledge of how to avoid HIV/AIDS ▪ Condom use at last and penultimate sex ▪ Relationship to last and penultimate sexual partner ▪ Length of time known last and penultimate sexual partner ▪ Age at first sex **Internet link:** http://www.measuredhs.com/
Multiple Indicator Cluster Survey (MICS), UNICEF	▪ Knowledge and source of knowledge of AIDS ▪ Knowledge of how to avoid HIV/AIDS ▪ Knowledge of testing sites and if ever tested **Internet link:** http://www.childinfo.org/index2.htm
Behavioural surveillance surveys (BSS), FHI	▪ Surveys for CSW, MSM, IDUs, Youth, Adults on HIV-related risk behaviours **Internet link:** http://www.fhi.org/en/topics/bss.htm

Discussing the table

Looking at Table 3.2 on the previous page, discuss the following questions:

a. Where can you find information about the age at first sex?

b. Which survey offers knowledge of testing sites?

Measurement Error

Measurement error occurs when the data collected do not accurately measure the characteristics of interest, which affects the validity of the data.

Measurement error terms

Measurement error is best illustrated with an example. Imagine the surveillance system wants to measure how often people have unsafe sex. To do this, the respondent is asked:

'Have you had unsafe sex in the last month?'

Measurement error could occur in one of two ways:

- The person had unsafe sex but is too embarrassed to admit it, so answers **'no**.**'** In this case, their behaviour is not measured accurately because the respondent does not tell the truth.

- The respondent thinks of unsafe sex as sex with the risk of pregnancy. In this case, we do not measure behaviour accurately because the respondent does not interpret the question as intended.

Some measurement error is more problematic than others. Suppose we are interested in measuring the average number of sexual partners in the last month for men and women, but in our imaginary population men are more likely to over-report the number of partners, and women to under-report the number of partners. Because of this over- and under-reporting, the difference between male and female behaviour appears larger than it is. This is *systematic error*. Systematic error can result in biased findings.

Measurement error terms, continued

Measurement error would be reduced if equal numbers of people over- and under-report their number of partners. Because the number of over-reporters is generally comparable to the number of under-reporters, the average number of partners calculated from the data will be the same as if everyone had told the truth.

However, the range of our measure will be larger than if everyone had told the truth. Because the over- and under-reporting balance each other out, this is *non-systematic error*. If measurement error is non-systematic, our finding will not be biased.

Note: *Bias* is a consequence of defects in the design or execution of a study. It cannot be controlled for during analysis or by increasing the sample size.

Systematic and non-systematic error example

Table 3.3 shows three measurement error scenarios for a sample of six people. The samples are described below the table.

Table 3.3. An example of measurement error.

	Scenario 1: True answers	Scenario 2: Non-systematic error	Scenario 3: Systematic error
Person 1	Yes	No	No
Person 2	No	Yes	No
Person 3	Yes	No	No
Person 4	No	Yes	No
Person 5	Yes	Yes	Yes
Person 6	No	No	No
% unsafe sex	50%	50%	17%

Scenario 1: All six people in the survey tell the truth. Three had unsafe sex and three did not. The true measure of unsafe sex in our sample is 50%.

Scenario 2: Persons 1, 2, 3 and 4 lie about having unsafe sex. Two people report that they had unsafe sex when they didn't. Two people report that they didn't have unsafe sex when they did. The error in reporting in not systematic (that is, in one direction) and our measure of unsafe sex in our sample is still 50%.

Systematic and non-systematic error example, continued

Scenario 3: Persons 1 and 3 lie about having unsafe sex. Both people report that they did not have unsafe sex when they did. The error in reporting is systematic (that is, all those who lie under-report unsafe sex) and our measure of unsafe sex in our sample is only 17%, so the measure is biased.

Discussing the table

Looking at Table 3.3 on the last page and the three scenarios, discuss the following questions:

a. How would the results change if only one person lied about reporting unsafe sex in scenario 3?

b. How can the % of unsafe sex be the same (50%) for scenarios 1 and 2?

Measurement error factors

In observational studies, measurement error can come from:

- questionnaire faults
- interviewer error
- respondent error

Questionnaire faults

The way that questions are phrased, set out, and ordered can affect how accurately the data collected reflects actual behaviour. Questionnaire faults include:

Culturally inappropriate questions - Some words and phrases may not be understood or may have different meanings in different cultures. For example, ‘unsafe sex’ may mean different things to different people. Such differences in meaning can result in respondents interpreting questions differently than how they were intended. This confusion may then result in confounding answers, and not accurately record the behaviour we are interested in. Some questions are understood, but considered objectionable in some cultures. This leads to untruthful answers. This can include questions of a personal nature or questions that convey negative implications.

Questionnaire faults, continued

Ambiguous wording - Vague questions produce vague answers. For example, '**Have you had sex without a condom recently?**' In this question, the term 'recently' is too subjective and 'sex' is an ambiguous term. However, questions can also be so precise that people cannot answer them.

Leading wording - Some questions are phrased in a way that makes the respondent more likely to choose a response category. The question may only mention some of the response options. For example, '**When should a condom be used: for oral sex, vaginal sex or what?**' The questions may also be leading, emotionally charged, threaten self esteem or indicate a socially desirable result. For example, '**You have had sex in the last year, haven't you?**'

Too many or demanding questions - Asking too many questions or questions that are too demanding can make respondents impatient or bored with the interview. An example of this would be asking a respondent to rank 25 factors in order of importance. In such a case, the respondent may stop listening carefully to the questions or provide answers they think will speed up the interview process.

Poor ordering of questions - Putting sensitive questions at the start of a questionnaire can put a respondent off and decrease the rapport between the interviewer and respondents. The respondent may stop listening carefully to the questions or provide answers that they think will speed up the interview process. Switching between topics or abruptly changing topic can also be confusing and cause respondents to lose concentration.

Interviewer error

The actions and behaviours of the interviewers can affect how accurately the data collected reflect actual behaviour. Interviewer errors include:

Not following the fieldwork protocols - Interviewers not following the fieldwork protocols can affect the accuracy of data. For example, in a quantitative survey, interviewers should ask the questions exactly as they are written on the questionnaire. This ensures that fieldworkers do not phrase or interpret questions in their own way. If fieldworkers do not follow this protocol, respondents may not be asked the same questions and measurement error may occur.

Interviewer error, continued

Being judgmental - If respondents feel that fieldworkers are judgmental and will disapprove of their behaviours, they may be less willing to report behaviours that are felt to be socially undesirable.

Misinterpretation of responses - Fieldworkers can misinterpret the responses that are given. For example, the respondent may respond using slang that the interviewer does not understand.

Mistakes in recording answers - Sometimes interviewers simply record the answers they are given incorrectly, or they may skip questions accidentally.

Respondent error

The actions and behaviours of the respondents can affect how accurately data collected reflect actual behaviour. Respondent errors include:

Misunderstanding the question

Faulty recall - Respondents can have difficulty remembering past behaviours correctly, or their memory of past behaviours can be influenced by behaviours or events that occurred more recently.

Untruthful answers - Respondents may give untruthful answers, especially if they perceive that there is a 'correct' answer. They may also over-report socially desirable behaviours and under-report socially undesirable behaviours. The personality, sex, race or age of the interviewer can influence the response.

Many questions essential to behavioural surveillance are sensitive. Sexual behaviour is difficult to study, owing to its private and sensitive nature. Also, doubts about the validity of reported behaviours abound. People are more likely to tell the truth in some situations than others.

Scepticism about the validity of behavioural data has been partially addressed by efforts to improve training, supervision and rapport-building with the survey populations. Results from sex surveys in the developed world are encouraging in their plausibility and consistency. Similarly, the data from similar surveys in low-income countries is now sufficiently accurate to merit serious analysis.

Data from large-scale surveys not specifically designed to study sexual behaviour may be less valid than data from surveys specially designed to measure sexual behaviour.

Reducing measurement error

Table 3.4 gives examples of how to each type of measurement error.

Table 3.4. How to reduce measurement error.

Source of measurement error	Possible solutions
Questionnaire faults	
Culturally inappropriate questions (additional information in Unit 7)	▪ Tailor wording of question to the type of respondent. ▪ Explore local terms with formative research and pre-test questionnaires. ▪ Place sensitive questions towards the end of the questionnaire. ▪ Make sure questions do not convey negative connotations. ▪ Soften the impact of sensitive questions by establishing a context. ▪ If translation is required, check by back translation.
Ambiguous wording	▪ Keep words a simple as possible; avoid jargon and abbreviations. ▪ Do not use vague questions or imprecise terms. ▪ Avoid presenting more than one concept at once. ▪ Make sure you do not assume too much knowledge.
Leading wording	▪ Mention all possible answers in the question or none. ▪ Do not use emotionally charged words and avoid threats to self-esteem.
Too many or demanding questions	▪ Be careful with questionnaire length; exclude questions you will not use. ▪ Use screening questions with skip patterns or not options that are not applicable. ▪ Avoid demanding questions and reduce demand with formatting.
Poor ordering of questions	▪ Place questions on the same topic together. ▪ Move from general to specific questions. ▪ Use transition statements when changing subjects. ▪ Place sensitive questions towards the end of the questionnaire.

Table 3.4. How to reduce measurement error, continued

Interviewer error	
Not following the fieldwork protocols	▪ Select and train fieldworkers well and produce a fieldworkers' manual. ▪ Ensure good supervision of fieldworkers. ▪ Ensure data quality check and procedures are put in place.
Misinterpretation of responses:	▪ Use interviewers who understand the survey population language. ▪ Select and train fieldworkers well and produce a fieldworkers' manual. ▪ Train fieldworkers to take notes whenever something is unclear.
Mistakes in recording answers	▪ Select and train fieldworkers well and produce a fieldworkers' manual. ▪ Ensure good supervision of fieldworkers. ▪ Use vertical answer formats and pre-code answers. ▪ Provide instructions for interviewers on the questionnaire. ▪ Use separate typefaces for instructions and answers.
Respondent error	
Faulty recall	▪ Ensure the time reference periods are appropriate.
Untruthful answers	▪ Use the appropriate data collection method. ▪ Place sensitive questions towards the end of the questionnaire. ▪ Do not use emotionally charged words and avoid threats to self-esteem. ▪ Use members of the target community as interviewers. ▪ Use interviewers of the appropriate age, sex and race.

Discussing the table

Looking at Table 3.4, discuss the following questions:

a. What are some ways to reduce ambiguous wording in a questionnaire?

b. How can you reduce interviewer mistakes in recording answers?

Importance of pre-testing

Pre-testing the questionnaire is an important step in developing appropriate survey instruments. Pre-testing is essential because it:

- determines the most appropriate sequencing of sections and questions
- identifies whether questions wording is clear and understandable for the interviewees and allows ambiguities to be identified and removed before the questionnaire is finalised

Importance of piloting

After the questionnaire is pre-tested, further improvements during piloting are required. Pilots usually consist of collecting data from small samples of respondents to check how well the planned fieldwork procedures work in practice. Piloting is essential because it:

- shows how long it takes to administer the questionnaires
- can inform the fieldworker training process by identifying sections or questions that cause problems, ideas and issues that are not clear, and skills that need to be strengthened
- helps plan fieldwork by showing how well fieldworkers cope with conditions
- identifying logistical constraints, and highlights potential problems with recording or checking data
- reveals factors that have not been considered during piloting

Fieldwork Practicalities

Working with hard-to-reach populations

High-risk groups that are included in surveillance are often hard to reach because:

- They engage in illegal/clandestine behaviours.
- They often do not want to be identified because of high levels of stigmatisation and discrimination.
- Their existence is denied by the general population and government.
- There are restrictions on who may approach the group and how the group can be approached (gatekeepers such as brothel owners may not want sex workers interviewed, the government may not want non-military personnel interviewing military, etc.).
- Group members have little time to talk.
- Groups do not want to be found for surveillance because they fear authorities or do not want outsiders entering the group.

Working with hard-to-reach populations, continued

- Group members are difficult to differentiate from non-group members. (For example, on entering a bar known to be a place where sex is sold, how do you differentiate a sex worker from a customer of the bar?)

Examples of practical difficulties and potential solutions of working with hard-to-reach groups are outlined in Table 3.5 below.

Table 3.5. Practical problems and solutions of working with hard to reach groups

Problem	Solution
How to identify group members	▪ Collect information on identification of members and in the places they congregate where they may be identifiable and accessible. ▪ Use members of the group as interviewers.
Fear of persecution decreases participation	▪ Conduct community entry activities. ▪ Work with authorities to ensure surveillance activities do not lead to police crackdowns, etc. ▪ Make sure there are some tangible benefits resulting from the surveillance activities. ▪ Use group members as interviewers or sensitise interviewers to issues. ▪ Work with NGOs who are accepted by the target population. ▪ Realise that it can take more than one surveillance round to gain a group's trust and persevere if there are problems in initial rounds.
Gatekeeper (person who controls access into the location) restriction	▪ Conduct community entry activities and explain the importance of surveillance and how they will ultimately benefit the population. ▪ Receive permission from the gatekeepers for the fieldwork. ▪ Involve gatekeepers in the surveillance activities. ▪ Work with NGOs who are accepted by the target population.
Scheduling	▪ Work around groups' schedule. ▪ Make interview as short as possible. ▪ Use incentives.
Consent and confidentiality	▪ Train fieldworkers on the importance of confidentiality and consent. ▪ Ensure that the gatekeepers know participation is voluntary. ▪ Identify potential problems with privacy and confidentiality.
Interviewer safety	▪ Select interviewers who know the area. ▪ Interviewers work in pairs.

Discussing the table

Looking at Table 3.5 on the previous page, discuss the following questions:

a. What is one method to reduce gatekeeper restrictions?

b. What are some methods to reduce interviewer safety concerns?

Selecting Appropriate Fieldworkers And Supervisors

Supervisor's role

The supervisor's role includes:

- representing the project at the community level
- negotiating with and accessing communities
- managing fieldwork and fieldwork teams
- organising logistics, travel, accommodation, and remuneration of fieldworkers
- supervising and supporting fieldworkers
- checking questionnaires and ensuring quality control
- exercise control over sampling
- ensuring ethical guidelines are followed
- dealing with problems/troubleshooting

It is common in research for supervisors to check data quality by re-interviewing a proportion of respondents. Through this process of quality control, it is ensured that:

- The interviewers conducted the interview correctly. That is, interviewers actually conducted the interview, didn't skip questions, recorded responses correctly and so forth.
- Respondents' answers are consistent. That is, do questions yield the same answer when re-asked? Are they reliable?

Re-interviewing can be difficult when conducting behavioural surveillance in hidden populations because names are not recorded on the questionnaire. Trying to relocate respondents is a breach of confidentiality.

Supervisors can instead check with people in the survey area to find out if the interviewers were there when they were supposed to be and how long they stayed. Supervisors can also perform spot checks where they re-interview the respondent straight after the interviewer has completed their interview. This makes relocating the respondent unnecessary.

Interviewer's role

The interviewer is responsible for collecting and recording good quality information from respondents on the questionnaires.

To fulfil these roles, supervisors and interviewers require specific characteristics and skills. Some of these skills can be taught through training. Others, such as fluency in appropriate languages, previous experience with research and with the target populations, educational considerations, gender, age and ethnicity considerations, cannot.

Summary

Measurement error occurs when the data collected do not accurately measure the characteristics of interest, which affects the validity. It is important to select appropriate data collection methods, since the type of data collection method can influence how respondents answer questions. Survey instruments that are widely used, reliable in their sampling procedures, statistical analysis, and are considered reliable as a source of behavioural data consist of the best design for behavioural surveillance.

Unit 3 Exercises

Warm-up review

Take a few minutes now to look back at your answers for the warm-up questions at the beginning of the unit. Make any changes you want to make.

Small group discussion

Get into small groups by country, region or province to discuss these questions.

1. In your experience, what are some examples of factors that affected measurement error in observational studies?

2. What are some general population surveys that you have used that ask questions about HIV/AIDS?

Apply what you've learned/ case study

Try this case study individually. We'll discuss the answers in class.

1. What topics need to be taught during interviewer training?

2. What topics need to be taught during supervisor training? Outline a curriculum.

Notes

Unit 4

Sampling Approaches

Overview

What this unit is about

In this unit, you will learn the theoretical and practical knowledge required to discuss and select sampling options for behavioural surveillance.

Warm-up questions

1. The _____ is the group that meets a survey's measurement objective (for example, all commercial sex workers in a city).

   a. target population
   b. survey population

2. Is drawing names randomly out of a hat for sampling an example of probability sampling or non-probability sampling?

   a. probability sampling
   b. non-probability sampling

3. True or false? Non-probability sampling is prone to selection bias.

   True False

4. We can increase the precision (that is, decrease standard error) of our estimate by increasing the_____.

   a. sample size
   b. quality of interviewer training

5. What is an estimate of precision that can be used to construct a range of values within which the true population measure is likely to fall?

   a. standard error
   b. systematic sample

Warm-up questions, continued

6. Which of the following is not a type of sampling?

   a. stratified
   b. cluster
   c. respondent driven
   d. systematic
   e. salient

7. True or false? Parametric statistics assume simple random sampling when performing tests.

   True False

8. True or false? A cluster sample provides more precise estimates of indicators than a simple random sample of the same size.

   True False

9. Which of the following is NOT a way to overcome sampling challenges for behavioural surveillance?

   a. using different sampling strategies for different groups
   b. using convenience sampling where possible
   c. using conventional sampling methods in unconventional ways
   d. using sampling techniques such as respondent-driven sampling (RDS).

Introduction

What you will learn

By the end of this unit, you should be able to:

- outline the purpose of sampling
- understand key theoretical concepts in sampling
- understand the need for more complex sampling designs
- understand sampling issues and options for behavioural surveillance
- understand the criteria for choosing a sampling approach

Sampling Approach

Definitions

Sampling is a way to measure characteristics for a specified target population when sufficient time and resources to obtain information from every member of the target population is not available. For example, if the target behaviour is condom use among commercial sex workers in a capital city, it is not feasible to identify and question every sex worker about their behaviour. Instead, selecting a sample to question serves as a viable statistical tool.

The *target population* is the group that meets a survey's measurement objective (for example, the adult population in a country or all commercial sex workers in a city).

A sample is individuals for study from the population. Once an estimate of selected indicators is obtained, the people selected are then interviewed. There are many different ways of selecting these individuals, called sampling approaches.

Sampling results in more reliable data if done correctly, and saves resources that might otherwise be spread across an entire population.

The survey population is the target population modified to take into account practical considerations (for example, adults living in households or all commercial sex workers in a city over the age of 15, excluding those who are based at home).

What do we want from a sample?

We want the sample to provide:

- unbiased estimates
- precise estimates

Unbiased estimates

When a sample is taken, the sample itself is not of interest. What the sample tells us about our target population is the main concern. The aim of the sample is to provide *unbiased* estimates of indicators for the target population. Avoiding bias due to sampling requires a random/probability sample. This ensures that the sample is statistically similar (or can be adjusted to be similar) to the population it was drawn from.

In a *probability sample,* each person in the survey population has a known, non-zero probability of selection.

If the sample is not a probability sample, or a probability sample was planned but not implemented correctly, findings cannot be generalised beyond the sample. This is because it is unknown whether the subjects in the sample are different from those in the rest of the target population. Statistical tests are not accurate when performed on non-probability samples.

Probability sampling example

Imagine that the people in our class make up the population of a city. We want to find out how many people in the city practice unsafe sex. An example of a probability sample would be if we wrote each person's name on a piece of paper and pulled a sample of names randomly from a hat, like a lottery. This is called *simple random sampling*.

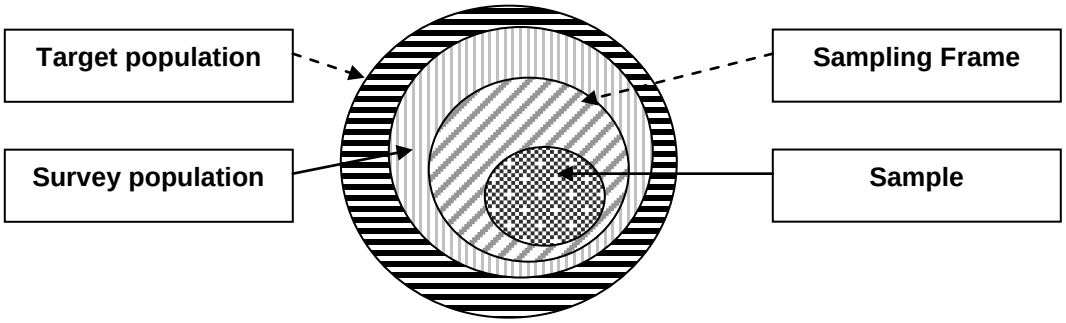
Everyone in the class has a chance of being sampled (a non-zero probability) and we know what that chance is. In simple random sampling, such as this, everyone has an equal chance of being selected. Therefore, if there were 10 people in the class, the probability of any individual being sampled would be 0.1 (1 out of 10). If there were 20 people in the class, the probability would be 0.05 (1 out of 20). Listing the names of the class members to draw from the hat is called developing a *sample frame*. A sample frame is a fundamental part of probability sampling.

Note: We could also have assigned each class member a number and selected numbers from a random number table.

A *sampling frame* is a list of units from which a sample may be selected. See Figure 4.1 next page to see the relationship among populations, sampling frame and the sample as a whole.

Probability sampling example, continued

Figure 4.1. The relationship among target population, survey population, sampling frame and sample.



Non-probability sampling example

An example of non-probability sampling would be if people from the class were not selected at random, but instead selected by the criteria of who arrived first this morning. This is easier to do than the simple random sample, as we do not have to develop a sampling frame (in this case, the list of people in the class).

The problem here is that the people who arrived first may not be the same as the rest of the class. They could be those who live closest to the class, those who own a car or those who don't have to get children ready for school before they come to class.

Whatever the reason, the early arrivals may be different to the rest of the class not just in their arrival time, but in the indicators we are interested in measuring. If we try to generalise what we find out about unsafe sex from these early birds, we could get a wrong (biased) idea about the sexual behaviour of the class as a whole.

In this type of sampling, members of the class did not have an equal probability sample. Those who arrive late have a zero probability (no chance) of being sampled. See Table 4.1 for a comparison of probability and non-probability sampling.

Non-probability sampling example, continued

Table 4.1. A summary of probability and non-probability sampling.

Issue	Probability sample	Non-probability sample
Prone to selection bias	No	Yes
Can generalise results to survey population	Yes	No
Can estimate precision of survey estimates (use statistical techniques accurately)	Yes	No
Usually requires sampling frame	Yes	No

Discussing the table

Looking at Table 4.1, discuss the following questions: If you need to generalise results to the survey population, which sampling should you use? Why is it generalisable?

Precise estimates

It is important that any given sample provide precise estimates. The information inferred from a sample only provides an estimate of the true population measure. There are many possible samples that could be selected from the population. Because of chance, each sample selected would produce a different estimate of our indicators. The variation in measurement that comes about by chance through sampling must be taken into account when using a sample to make inferences about a larger population.

When using probability sampling, how much our sample estimate may vary from the true population measure can be estimated. This estimation of precision is called the *standard error*. It can be used to construct a range of values within which the true population measure is likely to fall. A 95% confidence interval is the likelihood that the true population measure lies in our range, and what to aim for when taking a sample from a population.

One way that can increase the precision (that is, decrease standard error) of an estimate is increase the sample size. Standard equations are available to calculate required sample size, depending on:

- the initial starting level of the indicator
- the amount of change desired to be able to measure between surveillance rounds
- how certain the data will need to be able to detect change between surveillance rounds (the power of the study)
- the confidence interval necessary to determine that the true population measure falls within confidence limits (the precision of the study)

Simple Random Sampling Problems

Some of the drawbacks to simple random sampling include:

- It can require selecting a large number of random numbers from a random number table or names from a hat.
- Sampling frames for the whole population rarely exist and are too costly or impractical to compile.
- Populations can be spread over a wide area and the travel and time costs involved in covering the whole area can be prohibitive.
- The population consists of distinct and rare sub-groups of special interested.

Assigning random selection

Problem: Random sampling may require selecting a large number of random numbers from a random number table or names from a hat.

Solution: A more convenient method than having to pick names or numbers is to sample *systematically* from the sampling frame. In *systematic sampling,* we construct the sampling frame as in simple random sampling (that is, we make a list of everyone in the target population) but rather than selecting names or random numbers, we sample people at regular intervals down the list. For this scheme to work, select a random starting point. For this type of sampling to work, it must be ensured that the list is not ordered in any way that would bias who is selected in the survey.

No sampling frames for whole population

Problem: High cost or difficulty in creating a sampling frame for the whole population.

Solution: When it is difficult or impossible to make a list/sampling frame of each individual in the target population, we can develop a sampling frame of some larger unit. These are called *clusters* or *primary sampling units*. We then sample in stages by first sampling clusters and then sampling people within the clusters. Cluster sampling is the most common method of sampling in surveys, as it has the advantage that the sampling frame is not required to be a list of every person in the target population. Instead, a sampling frame of clusters is required. Once the clusters are selected, we are only required to list people in the selected clusters. All members of the target population still have a chance of being sampled (a non-zero probability) as long as all the clusters within which the target population is found are included in the list of clusters.

No sampling frames for whole population, continued

A *cluster* is any aggregate of the population of interest (for example, departments, villages, health facilities).

Clusters of unequal sizes need to be taken into account. If we do not take size into account, people in smaller clusters will have a higher probability of being selected than those in larger clusters. If people in smaller clusters are different to those in larger clusters, our sample population will not be like the target population. We may get a biased estimate of our indicators.

There are two main methods of ensuring that the sample provides unbiased estimates of the population parameter:

1. Make the probability that a cluster is sampled dependent on its size (PPS sampling). This requires that we know cluster size prior to cluster sampling.

2. Adjust for cluster size during data analysis. This requires that we have the correct statistical software to adjust for cluster size during the analysis.

Cluster sampling results in a less precise estimate of our indicators than simple random sampling, as respondents within clusters may be similar to each other. Compensate for this by increasing the sample size.

Populations spread out

Problem: Populations can be spread over a wide area and the travel and time costs involved in covering the whole area are prohibitive.

Solution: Use cluster sampling as described above, as it concentrates fieldwork in specific areas/clusters and reduces the fieldwork time and travel costs.

Distinct sub-groups

Problem: The population consists of distinct sub-groups of interest to surveillance.

Solution: When the population consists of distinct sub-groups, (for example, age groups or regions) make precise estimates of our indicators for each sub-group. If this is the case, we use *stratified* sampling.

Steps for stratification;

- Calculate the required sample size for measuring our indicator.
- Then define the sub-group (*strata*) and randomly sample the calculated sample size in each strata.

As we want to make precise estimates of our indicator for each strata, our sample size will be much larger than if we just wanted an estimate for the entire population. We can combine strata estimates to obtain a population estimate for our indicators. However, this requires that we know the proportion of the population in each strata.

Stratification is the classification of a survey population into sub-groups or strata on the basis of selected characteristics. *Stratified sampling* is the selection of separate (that is, independent) samples from each stratum. Stratification can result in a more precise estimate and is sometimes used purely to improve precision.

An example

The government of country Alpha desired to estimate the percentage of people in the general population who had participated in any unsafe sex in the last month. Alpha has three main regions: the coastal plain, the mountains and the semi-arid region. As the population is not evenly spread throughout the country, and it was thought that geography might influence the level of unsafe sex, the government chose stratified sampling. The sample size required to accurately estimate unsafe sex was calculated as 50 people. Alpha required estimates for unsafe sex in each strata/region, and needed to sample 50 people in each strata/region. If Alpha only wanted a countrywide estimate, their total sample size would be 50. Table 4.2 shows the results obtained with the estimated level of unsafe sex in each region.

An example, continued

Table 4.2. Results from a stratified sample carried out to estimate the level of unsafe sex in a country with three main geographical regions.

Region	Population	Sample size	Number had unsafe sex	% had unsafe sex	Estimated total who had unsafe sex
Coastal plain	1,500,000	50	30	60	900,000
Mountains	150,000	50	5	10	15,000
Semi-arid region	300,000	50	15	30	90,000
Total	1,950,000	150	60	33	1,005,000

Table 4.2 on the previous page shows how stratified sampling can be used to calculate the level of unsafe sex in each region. The level of unsafe sex for the population as a whole is calculated by estimating the total number of people in each region who had unsafe sex (for example, for the coastal plain this would be (60/100 * 1, 500, 000 = 900, 000), and then adding these up to get the estimated total number of people in the country who had unsafe sex (1, 005, 000). The total population size is 1 950 000, so the percentage of people in the whole population who had unsafe sex is 1,005,000/1,950,000 *100 = 52%.

You could calculate regional estimates without using stratified sampling during the analysis. However, there would be no guarantee that the sample size would be large enough to make precise estimates for each region.

Discussing the table

Looking at Table 4.2 on the previous page, discuss the following questions:

1. What percentage had unsafe sex in the semi-arid region?

2. What is the percentage of people in the whole population who had unsafe sex?

Sampling Issues in Behavioural Surveillance

Some sampling issues in behavioural surveillance include the following:

- Consistent sampling is required across survey rounds.
- General population surveys can rarely be used to access high-risk groups.
- Cluster sampling can be difficult when clusters are not stable.
- Members of high-risk groups may be difficult to identify and access.
- Cluster sampling is impossible if group members do not congregate.

Consistent sampling

Behavioural surveillance often aims to measure trends over time. Therefore, it is essential that the different survey rounds define and sample a sub-population consistently over time. If this is not done, it is difficult to know if any observed changes are real or just due to changes in methodology. Sometimes it is appropriate to change the sampling strategy because new and better techniques are developed. In this case, sampling the population using both old and new methods to compare estimates is the best practice.

Accessing high-risk groups

Household/general population surveys are rarely an appropriate method for locating members of high-risk groups. The group members may not be found in households in sufficient numbers through a household survey and may have behaviours that are too sensitive to discuss in a household setting. It is usually impossible to make a sampling frame of all the members of a high-risk group. One solution is to identify the places high-risk groups congregate, define these as clusters and sample these. Examples of possible clusters for high-risk groups are shown in Table 4.3 below.

Table 4.3. Examples of possible clusters.

High-risk group	Possible cluster
Brothel-based sex workers	Brothel
Non-brothel-based sex workers	Streets, bars, hotels, guesthouses
Men who have sex with men	Cruising sites
Intravenous drug users	Shooting galleries, injecting sites
Truckers	Loading/unloading/halting points
Migrants	Households, workplaces

Discussing the table

Looking at Table 4.3, discuss the following questions:

a. If you are unable to find a migrant worker in his or her household, where is another possible cluster? Why might this be sensitive information?

b. What are some possible clusters for intravenous drug users in your community?

Unstable clusters

Unless clusters are all the same size, a measure of cluster size is needed in order to ensure the sample is like the target population. It can be hard to measure the size of locations where high-risk groups congregate, as the individuals at the cluster are not fixed (for example, sex workers may move from one site to another). As well as the number of people, the type of people in a cluster may also vary (for example, sex workers who work in the afternoon may have different risk behaviours than sex workers who work in the evening). This makes it difficult to select a sample that is representative of the entire target population using conventional cluster sampling.

Hard to reach high-risk groups

High-risk groups can be hard to reach because members may be hidden and unwilling to be identified or acknowledge their risk behaviour. These difficulties have many implications for sampling:

- Constructing a sample frame of clusters can be difficult if people do not want to disclose the location (for example, brothels).
- Opposition from gatekeepers (those who control access into the location) may make including clusters in the sample problematic.
- Constructing sample frames within the selected clusters may be difficult, as individuals may not want to be identified as a member of the population.

High-risk groups do not congregate

Some high-risk groups do not congregate, making cluster sampling unfeasible. For example, it is difficult to think of a feasible cluster for home-based sex workers unless they all live in the same area. For other high-risk groups, only some of the population congregates. For example, it is possible to use cruising areas as clusters for some men who have sex with men. However, not all MSM frequent cruising areas, and an important section of the MSM population could be missed.

Potential solutions

There has been much debate about the best way to get around these sampling challenges. We need to:

- use different sampling strategies for different groups
- use conventional sampling methods in unconventional ways
- consider using sampling techniques such as respondent-driven sampling (RDS)

The adaptation of conventional sampling strategies and new techniques such as respondent-driven sampling will be discussed in the following section.

Sampling Options in Behavioural Surveillance

Conventional cluster sampling

For the reasons outlined above, conventional cluster sampling is not appropriate for many high-risk groups. However, it can be used for behavioural surveillance in the general population and youth. It can also be used for a few high-risk groups such as prisoners.

Time-location sampling

Time-location sampling (TLS) is like conventional cluster sampling, but it gets around the problem of clusters that are not stable (that is, clusters where the number and type of people varies by, for example, time of day). Time-location sampling allows the same site to be included in the sample frame more than once (for example, at different times of the day or different days of the week).

Thus, if the types of individuals in a cluster vary between weekday and weekends and between morning and afternoons, our clusters would be:

- Cluster 1= Site 1 weekday afternoon
- Cluster 2= Site 1 weekday evening
- Cluster 3= Site 1 weekend
- Cluster 4= Site 2 weekday afternoon
- Cluster 5= Site 2 weekday evening
- Cluster 6= Site 2 weekend, etc.

Once clusters have been selected, the most common approach is to randomly select the same number of respondents in each cluster and adjust the data during analysis for the fact that some location/time clusters will have more people associated with them than others.

This sampling scheme gets around the fact that the risk behaviour in a cluster may vary by time of day. It also means that it is not necessary to count the total number of individuals associated with a cluster, only the number of individuals in the sampling time interval. Measures are required to ensure that the same individuals are not interviewed more than once.

As for conventional cluster sampling, the sampling frame must cover the entire geographical universe of interest and include the majority of sites where group members congregate in sufficient numbers. Clusters should not consist solely of places that group members congregate for HIV prevention activities, as these locations are likely to be associated with people already concerned about HIV/AIDS.

Respondent-driven sampling

Respondent-driven sampling (RDS) is a sampling technique that does not require a sampling frame. It is an adaptation of a non-probability sampling method (snowball sampling) and is based on the assumption that members of the sub-population themselves can most efficiently identify and encourage the participation in surveillance of other sub-group members.

RDS starts with initial contacts, or seeds, who are surveyed. Each of these recruiters is given coupons to use to invite a fixed number of eligible people (often three) that he or she knows in the high-risk group being interviewed. Each new recruit brings his or her coupon to a central place and is interviewed. The recruit then becomes a recruiter. This occurs for five to six waves. Both the recruits and the recruiters are given incentives to encourage participation.

Theoretically, RDS should result in a probability sample. Given sufficiently long referral chains (five to six waves), the sample composition becomes stable, regardless of the people you started with, and the final sample will be like the population from which it is recruited. Computer packages exist to assist in the entry and analysis of RDS data to keep track of the:

- links between recruiters and recruits, so that we can calculate the probability of selection

- size of each individual's network so we can estimate how precisely the population measure is estimated by the sample estimate (to compensate for the fact the subjects are likely to recruit people like themselves and for difference in personal network size)

Table 4.4 on the next page is arranged with details about time-location sampling on the left and respondent-driven sampling on the right so that you may easily compare the steps for each method, as well as the advantages and disadvantages.

Table 4.4. The advantages and disadvantages of time-location and respondent-driven sampling.

Time-Location Sampling

Steps:

- Write protocol and calculate the required sample size.
- Identify clusters through *ethnographic mapping*.
- Construct a sample frame of clusters defining clusters by both location and time if the population is floating.
- Select clusters and individuals in clusters using equal probability sampling.

Advantages:

- Allows us to do a probability sample of populations that are hidden or floating.
- We only need a sample frame of clusters and individuals in selected clusters.
- Samples can be concentrated in geographical areas.

Disadvantages:

- Mapping and ethnographic work can be time-consuming and clusters/sites can close rapidly
- Only reaches subset of population that come into contact with the locations where the sampling is done
- It is difficult to identify and access respondents.
- It is difficult to maintain randomness while selecting respondents within clusters.
- PPS is not often done due to difficulties estimating cluster size. Samples often require weighting. This is not always done, resulting in biased estimates.

Respondent-Driven Sampling

Steps:

- Write protocol and calculate the required sample size.
- Start with initial contacts or ‘seeds’ who are surveyed and then become recruiters.
- Each recruiter invites up to three people they know in the high-risk group to be interviewed.
- The new recruits become the recruiters.
- Five to six recruitment waves occur.

Advantages:

- Field operations are not difficult.
- There is much less need for *ethnographic mapping* or a sampling frame.
- The target population recruits for you. This is good when the group does not trust the research community.
- Less visible members of the population are reached.
- The cost is lower than for time-location sampling.

Disadvantages:

- The population must be a network.
- You need to keep track of links between recruiters and recruits.
- Ethical issues are involved in using incentives.
- RDSAT, the statistical software used for RDS studies, is functional but somewhat limited.

Discussing the table

Looking at Table 4.4, discuss the following questions:

a. How many recruitment waves occur in RDS?

b. What are some of the advantages of using TLS?

c. Which type of sampling, RDS or TLS, would be preferable in for a study of sex workers in your country if sex work was illegal? Why?

d. Would your answer be the same for answer c if sex work is legal and your study concerns sex workers in brothels? Explain.

Sampling approach criteria

You should assess sampling options for each sub-population of interest. Answering the following questions can help guide the selection of sampling strategies:

1. Is the population of interest the general population or youth? If **yes**, conventional cluster sampling is recommended.

2. Do group members congregate in accessible locations/sites in high proportions? If **no**, RDS is recommended.

3. Is it possible to construct a list of all group members associated with each site? If **no**, TLS or RDS is recommended.

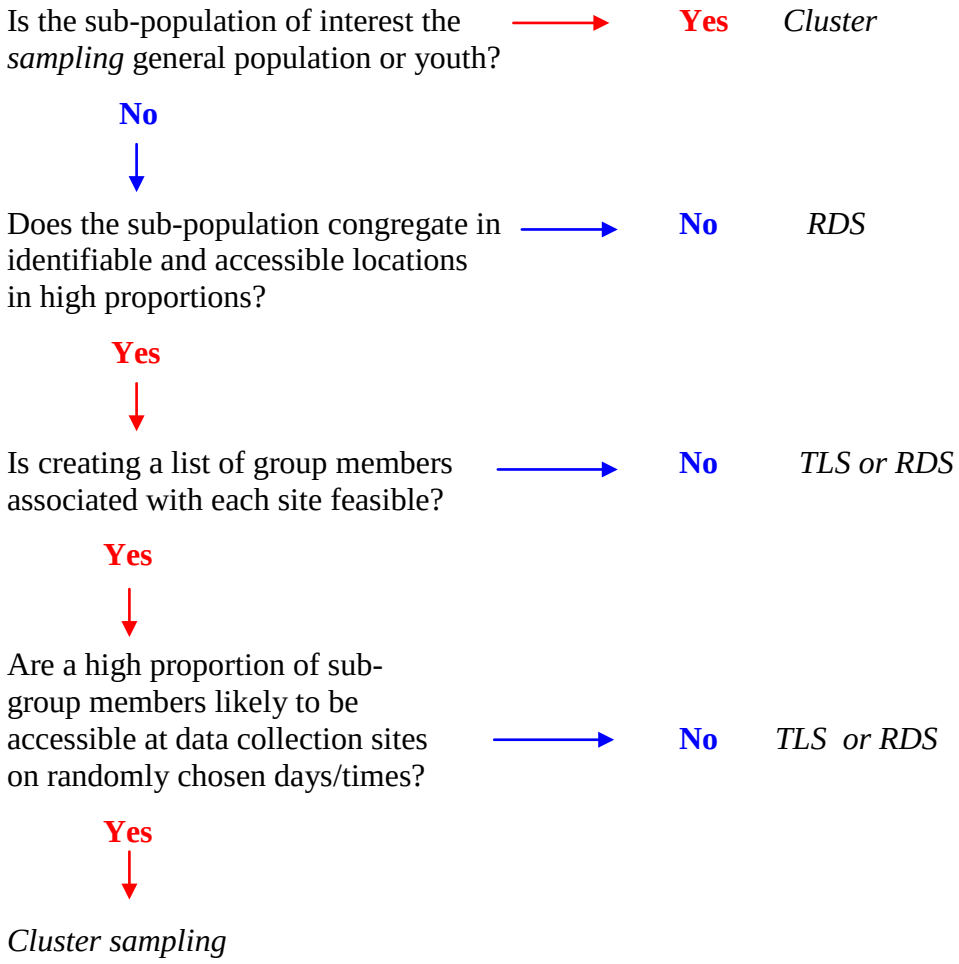
4. Are all group members on the list (not just those who happen to be present at a site) readily accessible during data collection? If **no**, TLS or RDS is recommended.

When answers to questions two through four are “yes,” (as well as for the general population and youth), conventional household or institutional survey methods are preferable.

These questions are represented diagrammatically on the next page.

Sampling approach criteria, continued

Figure 4.1. Choosing a sampling approach.



Sample Size Calculation

Formula for Sample size Calculation

The sample size needed to conduct behavioural surveys can be based on the number of participants needed in each round (or year) to detect a change in the proportion of an indicator from one round to the next. For example, you would like enough sex workers in your survey rounds to show that condom use during a population's last paid sex act increased from 20% in the year 2006 to 30% in 2007.

The general formula for the needed sample size (n) is:

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

Where:

n	=	Sample size required per survey round (year).
D	=	Design effect (see below)
$Z_{1-\alpha}$	=	The z score for the desired confidence level, usually 1.96 for 95%
$Z_{1-\beta}$	=	The z score for the desired power, usually 0.83 for 80%
P_1	=	The proportion of the sample reporting indicator in year 1
P_2	=	The proportion of the sample reporting indicator in year 2
P	=	$(P_1 + P_2)/2$

Choosing the values of these numbers is based on the following considerations.

D design effect. The design effect can be thought of as a correction factor for how much a cluster sample differs from a simple random sample. Effectively, the design effect multiplies the sample size by the factor of D. The design effect accounts for the similarities people have when they are sampled within the same cluster. For example, female sex workers within a particular brothel may be similar with respect to condom use because of the social norms, condom availability, or intervention programmes of the particular brothel. Choosing a design effect is difficult without prior survey data. Design effects from 1 (i.e., none) to 2 (moderate) cover a typical range.

Formula for sample size calculation, continued

For RDS surveys, a design effect of 2.0 is recommended. For cluster sampling and TLS a design effect of 2.0 is also recommended. ***The bigger the D, the larger the sample size needed.***

P_1 and P_2. P_1 and P_2 are the measures of interest for which you wish to see a change between survey rounds. For example, you wish to show that condom use at last paid sex act for sex workers increased from 20% in 2006 (P_1) to 30% or greater in 2007 (P_2). P_1 is usually based on previous surveys in the same or similar population, or an educated guess at what the level will be. P_2 is ideally set at the goal you would like to achieve (e.g., a 10% or greater increase in condom use). In practice, it is usually set at the smallest change you think is meaningful. For example, a 10% increase in condom use would be considered a meaningful improvement, whereas a 1% increase would not be considered meaningful.

The smaller the change you wish to detect, the larger the sample size you will need. Also, the closer P_1 and P_2 are to 50%, the larger the sample size you will need.

$Z_{1-\alpha}$. The $Z_{1-\alpha}$ score is a statistic that corresponds to the level of significance desired. Usually, a significance level of 0.05 (or, equivalently, a 95% confidence level) is selected and corresponds to a value of 1.96. This value is used when the change in the indicator might be either up (increase) or down (decrease) from year to year (a "two-tailed" statistic). ***The smaller the significance level (i.e., higher confidence level), the larger the sample size you will need.***

$Z_{1-\beta}$. The $Z_{1-\beta}$ score is a statistic that corresponds to the power desired. Usually, 80% power is selected and corresponds to a value of 0.83. This value is used when the change in the indicator might be either up (increase) or down (decrease) from year to year (a "two-tailed" statistic). ***The higher the power, the larger the sample size you will need.***

Example of sample size calculation

Suppose you are planning a survey of sex workers using a two-stage cluster design. You wish to ascertain whether condom use will increase from 20% in the baseline survey (this year) to 30% or greater in the survey wave next year. How many sex workers do you need to include each year?

Solution:

D = 2 (moderate)
$Z_{1-\alpha}$ = 1.96 (95% confidence level)
$Z_{1-\beta}$ = 0.83 (80% power)
P_1 = 20% condom use in year 1
P_2 = 30% condom use in year 2
P = (.20 + .30)/2 = .25

$$N = 2\{1.96\ SQT[2x.25(1 - .25)] + 0.83\ SQT[.20(1-.20) + .30(1-.3))]\}^2/(.30 - .20)^2$$
= 582 sex workers per survey wave

Table 4.5 provides pre-calculated sample size estimates over a range of possible scenarios in behavioural surveillance.

Table 4.5. Sample size needed per survey wave to detect a change in the proportion of an indicator between survey waves using a 95% confidence level, 80% power and a design effect of 2.0.

Indicator level (P1) at time 1	Indicator level in time 2 (P2)	Sample size needed each year with a design effect of 2.0
.10	.20	395
.10	.25	197
.20	.30	581
.20	.35	274
.30	.40	706
.30	.45	322
.40	.50	768
.40	.55	343
.50	.60	768
.50	.65	336
.60	.70	706
.60	.75	301
.70	.80	581
.70	.85	239
.80	.90	395
.80	.95	149

Discussing the table

Using Table 4.5, find the sample size you would need to do an RDS survey of IDU in your area. Justify your answer in terms of which indicator you are using and what the level of the indicator will be in wave 1 and wave 2.

Summary

We sample when we desire to measure characteristics for a specified target population but lack the time and resources to obtain information from every member of the target population. Concentrating survey time and resources on questioning a sample of people can also result in better quality data than spreading resources over the entire population. It is also important to understand sampling issues and options for behavioural surveillance. You should also follow the criteria for choosing the most appropriate sampling approach.

Unit 4 Exercises

Warm-up review

Take a few minutes now to look back at your answers for the warm-up questions at the beginning of the unit. Make any changes you want to make.

Small group discussion

Get into small groups by country, region or province to discuss this question.

a. What sampling strategies have you had experience with?

b. What difficulties and successes did you have with the strategy?

Apply what you've learned/ case study

Try this case study individually. We'll discuss the answer in class.

1. For each of the following groups, decide:
   - What is the best sampling strategy
   - Why this is the best strategy?
   - What are the strong and weak points of using this method for the group?

   a. Group 1: Youth

   b. Group 2: MSM

2. Using Table 4.5, find the sample size you would need to do a respondent-driven sampling (RDS) survey of MSM. A survey of MSM in your city found that 80% of MSM could correctly name the ways that HIV is transmitted and ways that HIV can be prevented. Health education campaigns are planned for the coming year using MSM peer educators. You wish to show that HIV/AIDS knowledge improved by at least 10% by the following year. How many MSM do you need to include in this year's survey and next year's survey? Assume that knowledge may increase or decrease and you wish to have 80% power. What sample size would you need if you decided to use time-location sampling (TLS) instead of RDS?

Notes

Unit 5

Data Analysis and Use

Overview

What this unit is about

In this unit, you will learn the skills required to ensure behavioural surveillance data is analysed, disseminated and used appropriately.

Warm-up questions

1. True or false? It is better to share over-all responsibility between different staff for the data management.

   True False

2. Data management does not include which of the following?

   a. data coding
   b. data entry
   c. data cleaning and checking
   d. data framing

3. True or false? The data manager should not be involved during the questionnaire design.

   True False

4. Which type of behavioural surveillance analysis is performed to determine whether one variable is related to the distribution of another (for example, an association between a respondent's age and their use of condoms)?

   a. univarite
   b. bivariate
   c. multivariate

Warm-up questions, continued

5. Most of the indicators defined for behavioural surveillance purposes are calculated through _____ analysis.

   a. univarite
   b. bivariate
   c. multivariate

6. When data are presented, package appropriately for the _____.

   a. different audience
   b. data collection sequence

7. True or false? The surveillance cycle ends when the official report is published.

   True False

8. True or false? Data from other sources, including other countries, can be used to fill in any important gaps as long as the source is made very clear.

   True False

Introduction

What you will learn

By the end of this unit, you should be able to:

- discuss data management issues
- describe the types of data analysis commonly used in behavioural surveillance
- understand the steps in ensuring appropriate data analysis and use
- list the different audiences for behavioural surveillance data
- interpret and package data appropriately for the different audiences

Data Management Issues and Activities

What is data management?

Collecting data is expensive, so efficient and effective data management is critical. Data management includes:

- data coding
- data entry
- keeping the data safe
- cleaning and checking the data
- preparing the data ready for analysis (merging, coding, etc.)

These activities cost time and money, and can affect the quality and usefulness of the data. Data management should be considered prior to data collection. This is when it can be useful to establish simple data management guidelines, covering issues including coding, storing questionnaires, data entry, and database management. These guidelines should be followed from the start, and deviations should be documented. Data management should be included on the agenda of surveillance meetings, so that it is considered by the whole surveillance team.

Issues to consider in data management

Personnel: Data managers with experience and skills are required to manage large and complex datasets. It is usually better for one person to have over-all responsibility for the data than to share responsibility between different staff. The data manager should be involved during the questionnaire design stage, not only to review the questions, but to review the way in which the questionnaire is organised. Make sure that data entry clerks and interviewers receive adequate training to minimise errors. You should also make sure that their work is monitored carefully.

Issues to consider in data management, continued

Computer capacity*:* Analyse likely computer capacity requirements carefully. Databases often take up more space than anticipated, and it is important to allow enough capacity for storage of data.

Back-up routine: Back up regularly to prevent loss of some or all of your data files.

Audit trail: Document everything you do and when you do it (for example, data entry, verification, corrections). Include notes about any queries. Recording changes to the master dataset on paper provides an audit trail. You can also have an electronic logging system that records all changes.

Data checking: Data errors can be introduced at any stage of the surveillance process and it is therefore important to be vigilant throughout. Data checks should include:

- consistency checks by supervisors after data collection
- double data entry, as it is unlikely that two people will make the same mistake
- data-entry programmes that can check that values stay within specified realistic ranges
- programmes written specifically to check for inconsistencies in the data, so any errors found can be checked back to the questionnaires and, if necessary, the fieldworkers

Database merging: With several data entry clerks working on the data, and several copies of the database, you need a method for merging the data at the end of the data entry phase.

Management of paper questionnaires: Take care of the questionnaires, making sure they are protected from the elements and kept confidential. Signing questionnaires at each stage of the data collection (data checking) is a useful way to keep track of which data should be on the database and which are still to be entered.

Storage and archiving: Develop a system for archiving raw data. The data management system should document data effectively, allow you to find the data you need quickly, and keep data archives and back-ups safe, up-to-date and usable.

Data Analysis

What is data analysis?

Data analysis is summarising, presenting and interpreting data. Behavioural surveillance analysis can either be:

- *cross-sectional* - analysing data from one surveillance round
- *trend* - analysing data over several surveillance rounds.

If probability sampling was used, statistical tests can tell us how well the sample measure estimates the true population measure. That is, they can tell us how likely it is that our findings could have occurred by chance sampling variation or whether they really represent true population values. You need special analytical techniques if the sample was a cluster sample or if the cluster sample did not use a self-weighted design.

Types of data analysis

There are three main types of analysis conducted in behavioural surveillance:

- Univariate analysis: This is the most basic type of behavioural surveillance analysis. However, it is often the most important, because it shows the distribution of each variable. Most of the indicators defined for behavioural surveillance purposes are calculated through univariate analysis. They would include variables like the proportion of young men who have had sex with more than one partner during a given time period. When trends are analysed, statistical techniques are used to calculate how likely it is that changes in the proportions could have occurred by chance, or whether observed changes are likely to reflect real changes.

- Bivariate analysis: This analysis is performed to determine whether one variable is related to the distribution of another. For example, there might be an association between a respondent's age (the explanatory variable) and their use of condoms (the outcome variable). Variables are associated if the value of one tells you something about the value of another. Statistical tests in bivariate analysis determine whether any observed difference reflects a true difference, or may be due to chance.

- Multivariate analysis: This analysis is performed to look at the influence of at least two variables on another variable, since relationships between variables are often complex and interwoven. Multivariate techniques can pinpoint the individual effects of several explanatory variables on an outcome variable that may be related to each other.

Steps in data analysis and use

Follow these four steps when performing data analysis.

1. Develop an analysis plan.
2. Explore the data.
3. Use appropriate statistical techniques.
4. Interpret data.

Develop an analysis plan

Before data collection begins, an analysis plan should be produced. This ensures that the correct data are collected, and that sufficient time and resources are allocated to data analysis. The plan should include:

- Listing the questions surveillance needs to answer. This should be done in consultation with the final data users to ensure their data needs are met.
- Defining the indicators needed to answer these questions.
- Making mock-ups of key tables and graphs to ensure you know how all the data collected will be used.
- Identifying personnel with appropriate skills, software needs, etc. It can be important to involve statisticians from the start. If they understand the purpose of the research and are familiar with the questionnaires, they will provide more useful inputs than if they are just brought in at the end to do multivariate analysis.
- Allocating sufficient funds and time for analysis. Cross-checking, recoding and the creation of indicators is often what takes time rather than doing the actual analysis.

Explore the data

Data analysis itself is usually done by a trained statistician. Once the data have been collected and cleaned, the statistician should first explore the data so they understand the data coding and detect any errors in the data. This should be followed by producing frequency distributions of the variables and recoding the data to create the required indictors.

Use appropriate techniques

When the descriptive summaries have been carried out, the statistician should look carefully at the results (such as checking for consistency in sample sizes and missing values) for errors. These results will focus further analysis. Most behavioural surveillance analysis will be univariate and bivariate, but whatever analysis is used, it should start with simple analyses that have a broad focus. Key findings can be missed if analysis immediately focuses on details.

Data weighting and multivariate analysis add a layer of high science to analysis. Nevertheless, significant room for error remains and complex analysis is not always appropriate, particularly if data quality is poor or sample sizes are insufficient.

Interpret data

Surveillance data, like any other data, can often be interpreted in more than one way. How indicators are defined and the selective presentation of indicators can greatly affect how the data are interpreted. People are very often driven by their own personal bias to give greater weight to one interpretation over another. In addition, sometimes people miss the whole story simply because they focus too closely on the details. People look at pieces of data as separate entities rather than as different aspects of the same story. The importance of examining information on population size, HIV prevalence, STI prevalence and risk behaviour together cannot be overemphasised. We will use the data in Table 5.1 on the next page to explore issues around data interpretation.

The validity of self-reported data on sexual behaviour and drug use is frequently questionable. This does not mean that surveillance data are worthless. One should merely be cautious about jumping to quick conclusions about small rises or falls in infection, and surprising findings in behavioural surveillance should be investigated further.

Table 5.1: Results of surveys on sexual behaviour among high school students (1998-2002)*.

Year	Had sex in past year (% of all)	Had multiple partners in the last year (% of all)	Always used condoms (% of those with multiple partners)
1998	73	9	50
1999	70	11	48
2000	68	15	41
2001	58	20	38
2002	48	24	35

*Abstinence-only sexual health education curriculum was introduced in late 1998 (three months after the first surveillance round).

Discussing the table

Looking at Table 5.1, discuss the following questions:

1. Based on columns 3 and 4, could you argue that abstinence programs are not working?

2. How do the percentages of those who had sex in the past year compare between 1998 and 2002?

Using Behavioural Surveillance Data

Steps in data use

Currently, in most surveillance systems a gap remains between the collection and the use of data. One reason data are not better used is that surveillance systems are often fragmented. When no single entity is responsible for compiling, analysing and presenting data cohesively, the different groups involved in surveillance often consider their job done. Before data are used, form a surveillance committee to help resolve this issue.

There are two steps to follow in using behavioural surveillance data:

- Develop data use plan
- Present data

Develop data use plan

Before data collection begins, produce a data-use plan. This helps ensure that information gets to the right people in a timely fashion and in a way they understand.

Developing this plan should be tied to identifying the country's data needs and consequently the data audience (that is, who will use the surveillance data).

At this stage it can be useful to identify surveillance outputs for the different audiences and how they will be disseminated. These could include a national report on HIV/AIDS, policy briefs, press releases, etc. Once outputs are identified, sufficient funds and time can be allocated for these activities. If possible, networks for future advocacy should start being built before you go ahead with surveillance activities. It is very important to establish good relationships with policy-makers and other data users.

Some common data users include:

- politicians and policy-makers
- HIV/AIDS programme managers in health and other sectors
- international agencies
- NGOs
- researchers
- populations included in the surveillance system
- surveillance system personnel
- the private sector
- the press
- legal professionals

Present data appropriately and accurately

Presenting data appropriately and accurately helps ensure it is used. When data are presented, consider the following issues.

- Packaged appropriately for the different audiences: The different audiences will respond to the data presented in different ways. For example, face-to-face meetings, videos or briefing sessions are often more effective ways to interest decision-makers than producing lengthy documents that they do not have time to read. Choosing the right product for the right audience will be covered in detail later in this unit.

Present data appropriately and accurately, continued

- Be realistic about the limitations of the data (sample size, response rates, threats to quality, etc.): Surveillance data does need to be perfect. They just need to be good enough to give a reliable idea of the major trends in HIV infection and related risk behaviour. Once numbers get presented as stand-alone facts, the information that would help people gauge the quality of the data is often lost. Limitations in the data in terms of sample size, response rates, etc. need to be reported and the data should not be presented as more "scientific" than is possible in the context. For example, surveillance data is sometimes presented to two decimal points of accuracy, giving an air of statistical solidity, when presenting rounded numbers is a more accurate reflection of data quality.

- Use all data sources available: Surveillance systems rarely produce all of the data that meets local advocacy needs. Data from other sources, including other countries, can be used to fill in any important gaps as long as the source is made very clear.

- Take care with the physical presentation of the data: Most people who work in public health can look at a table or graph and understand what it means, but many people in other fields cannot. If needed, headlines should be used to tell people what the data mean rather than simply describing the data. Generally, graphics should be kept as simple as possible and free of clutter.

Choose the right product for your audience

Public health officials often take one-size-fits-all approach to data use, believing that surveillance work ends when the official surveillance report is published. The report may indeed contain all the important information about the levels and trends of HIV and risk behaviours. However, the same data need to be presented differently for different audiences to be able to sell the key messages and get the data acted upon.

Successful advocacy follows a number of rules:

- Define your goals
- Define your audience
- Find out what influences their thinking and how to get their attention
- Use the right language
- Get the length right
- Choose the best messenger
- Get the timing right

Define your goals

Surveillance data can meet several goals and data needs. Each goal may have to be presented to different audiences in different languages.

Define your audience

With whom must the public health official communicate to ensure the goals defined are obtained? Who has the potential to push things forward and who could potentially obstruct progress?

Find out how to get their attention

Once the audiences have been identified, you must determine the best way of communicating with them and the key messages for the different audiences. The best communication occurs when people have something in common. Therefore, understanding the concerns, motivation and objectives of each audience is important. Once you understand their concerns, try to use the data to address these.

Use the right language

After you identify the goal, audience, key messages that will appeal to them and the data used to make the case, language is important. For example, will confidence intervals mean anything to the group or will they confuse the picture?

Get the length right

A report or presentation has no value if no one reads it or listens. This means fitting the key information into the time that people are prepared to dedicate to it. If HIV/AIDS is the core interest of your audience, they may appreciate a full report they can read through and keep for their reference. If HIV/AIDS is not a core area of interest, the audience is more likely to digest the information if it is presented in a one-page fact sheet or short brochure.

Choose the best messenger

People listen to the people they trust. For example, the minister of health may be the best person to present surveillance data at cabinet level, whereas young people may pay more attention to a pop star acting as an AIDS ambassador.

Get the timing right

HIV/AIDS is not the only issue on people's agendas. One way of increasing the attention the message gets is to time the release appropriately. For example, avoid clashing with important events and make use of events already scheduled, such as World AIDS Day.

Note: *Guidelines for the Effective Use of Data from HIV Surveillance Systems (2004)* elaborates on how to present data for different audiences and provides good examples of how HIV data can be used.

Summary

Since there are several data management issues, efficient and effective data management is critical. Behavioural surveillance data analysis can either be cross-sectional or trend. IT is important to understand the steps in ensuring appropriate data analysis. You should also understand the audiences for behavioural surveillance data and how to present and package data appropriately for the different audiences.

Unit 5 Exercises

Warm-up review

Take a few minutes now to look back at your answers for the warm-up questions at the beginning of the unit. Make any changes you want to make.

Small group discussion

Get into small groups by country, region or province to discuss these questions:

1. Discuss the strengths and weaknesses of the data management, data analysis and data use system used for behavioural surveillance in your setting.

2. Use data from your country's annual report. Write a paragraph for the general public, describing the data.

3. Design a flyer for the general public that conveys the data in Question 2 to the general public.

Apply what you've learned/ case study

Try this case study individually.

How do you think data would be best presented to a senior politician in terms of language, length and messenger?

Notes

Unit 6

Ethical Considerations

Overview

What this unit is about

In this unit, you will learn about ethical considerations and requirements in behavioural surveillance.

Warm-up questions

1. Match each ethical principle with its definition.

____	Respect for persons	a. Refers to minimising risk to individuals—not only physical risk, but also risk of psychological harm and stigmatisation.
____	Beneficence	b. Requires investigators to see study subjects not as passive sources of data, but as persons whose rights and welfare must be protected.
____	Justice	c. Risks and benefits from studies should be distributed fairly and evenly in populations.

2. It is important that everyone the surveillance system encounters is treated with_____, from community leaders and local officials to those surveyed.

3. What is informed consent?

4. Name two pieces of information that should be provided before a person can make an informed decision to take part in a survey?

   a.

   b.

Warm-up questions, continued

5. True or false? In surveillance, written consent forms are the most appropriate way to document that the process of informed consent has occurred.

   True False

6. What are two ways to help ensure a participant's confidentiality?

   a.

   b.

7. Match each of the following ethical issues with a potential solution:

____	Loss of earnings	a. Ensure fully informed consent.
____	Increases discrimination of the group	b. Conduct interviews outside work times.
____	Participants get no direct benefit from surveillance	c. Do not foster false expectations.

Introduction

What you will learn

By the end of this unit, you should be able to:

- understand the basic ethical principles of working with humans
- define informed consent and the procedures that help ensure it
- understand the importance of confidentiality and how to ensure it
- discuss the ethical consideration unique to behavioural surveillance

Ethical Principles of Working With Humans

Three universally accepted ethical principles of working with humans are *respect for persons*, *beneficence* and *justice*:

- Respect for persons - requires investigators to see study subjects not as passive sources of data but persons whose rights and welfare must be protected.

- Beneficence - refers to minimising risk to individuals—not only physical risk but also risk of psychological harm and stigmatisation.

- Justice - means that risks and benefits from studies should be distributed fairly and evenly in populations.

It is important that everyone the surveillance system encounters is treated with respect, from community leaders and local officials to those surveyed. This is important in every aspect of the study from engaging people in the study, to the way interviews are conducted, and in the importance of providing feedback to the participants and communities involved.

Informed Consent

Before being interviewed, respondents need to decide whether they want to participate in the survey or not. *Informed consent* means that you tell the person enough about the nature of the surveillance for them to make a proper (informed) decision about whether or not to take part.

No project staff should pressure, coerce or deceive respondents in an effort to ensure their participation and staff should also try to ensure that respondents are not pressured by other family or community members. Ensuring that the decision is not influenced may involve discussions around the voluntary nature of participation with gatekeepers and including overviews of informed consent in the fieldworker training and manual.

Information that should be provided

A person can only make an informed decision to take part in a survey (give their consent) when he or she has information about the procedures and purpose of the interview. During informed consent, subjects are provided with information about the survey, given an opportunity to ask questions and then given the opportunity to decide whether to participate or not. This information is usually provided on an information sheet, which the fieldworker reads to the participant. Studies have shown that getting voluntary and informed consent in developing countries is difficult. You must make sure that the information sheet uses appropriate language, is not so long that participants stop listening, and contains all the essential information.

Information that is relevant to the subject's decision on whether or not to participate includes:

- the nature of the survey (for example, who is conducting the survey, purpose of the survey, length of interview, type of questions, etc.)
- the potential risks and benefits
- how the information will be used
- how their privacy will be protected (names or addresses are not written)
- that participation is voluntary
- that participants have the right to refuse to answer any questions or stop the interview at any time, especially as they may find some of the questions sensitive

Documenting informed consent

Written consent forms are generally required to document that the process of informed consent has occurred. In surveillance, in order to ensure total confidentiality, it is usually best to obtain verbal consent. This means that the name of the respondent does not need to be recorded. There still needs to be some way of verifying consent, but rather than the participant's signature, interviewers can sign a statement to verify that the respondent has been given the required information and has decided to participate.

Maximising participation

Although consent must be voluntary, we want to try and maximise participation to reduce bias. Methods include keeping interviews as short as possible, conducting fieldwork at times that are convenient to the participants and stressing the altruistic benefits of participating. Reducing refusals is important because those who refuse to be interviewed may be different from those who participate, and if there are lots of refusals, our sample may not be like the population of interest and may be biased. We need to facilitate participation, but without being coercive.

Response rates should always be reported in the analysis. Participation bias should be assessed and taken into consideration in the analysis. Recording the following for each cluster can be useful for this assessment:

- number of target group members present at the cluster at the time of interview
- number interviewed
- number who refused to be interviewed, their reason for refusing and basic socio-demographic information about them
- number that were not interviewed for other reasons
- number rejected as duplicates

Using incentives

Incentives can consist of cash payments for participation or small gifts, such as T-shirts. In general, incentives are considered appropriate for compensating or thanking study participants for time away from work and out-of-pocket expenses, such as transportation. However, higher payments may jeopardise the voluntary nature of informed consent. They can create a situation where an individual's decision to participate is unduly influenced by money or gifts.

Additionally, using incentives may result in a sample that is not like the population of interest because the sample is biased towards those who have a greater need for the incentive. This needs to be balanced against the fact that not using incentives may bias the sample towards those who are more cooperative.

Respondent-driven sampling provides incentives to participants to recruit additional members of the high-risk population to the study. These incentives can be considered 'payment' to the participant, who in their role as recruiters act as fieldworkers. This part of the methodology may be controversial in some settings, and may require explanation to the institutional review board reviewing and overseeing the study.

Confidentiality

Confidentiality protects subjects from adverse consequences that may arise from other people knowing their responses or that they participated. For example, if information about a person's sexual preference is disclosed, he or she may suffer discrimination, stigma or even be subject to criminal charges. Potential threats to confidentiality, as well as measures taken to minimise them, should be discussed with the participants as part of the informed consent process. The main ways to ensure confidentiality include:

- ensure names or other means of identification are not recorded on surveys
- store data safely and appropriately
- train fieldworkers on the importance of confidentiality
- have clear disciplinary procedures for staff who breach confidentially
- identify problems and possible solutions related to confidentiality

Threats to confidentiality

Threats to confidentiality include finding a private place to conduct interviews and stopping other persons or gatekeepers from being present during the interview.

The presence of other people breaches confidentiality and may cause the respondent embarrassment and influence some of his/her answers. In such cases, the interviewer can explain to the respondents that some questions are confidential and ask them to suggest a place where you are unlikely to be disturbed.

Sometimes fieldworkers hear stories during interviews that make them so sad that they need to talk about it. Fieldworker training should stress that although it is important to talk through depressing issues, this should only be done with other team members and in a way that does not easily identify the respondent.

Note: Issues around maintaining confidentiality are more complicated when surveillance wishes to link individuals behavioural and biological surveillance data.

Ethical Considerations Unique to Behavioural Surveillance

Working with high-risk groups

Some high-risk groups, such as MSM, CSWs, and IDUs, participate in illegal or stigmatised behaviour and there is the potential for surveillance to make this stigmatisation worse.

A list of potential harms is listed in Table 6.1 below.

Table 6.1. Potential harms caused by behavioural surveillance in high-risk groups.

Potential harms caused by behavioural surveillance in high-risk groups
▪ Physical (public attack, abuse, loss of healthcare services) ▪ Legal (arrest, prosecution) ▪ Social (disclosure to family, workplace, discrimination, loss of employment, isolation)

If high-risk group members fear that information about their behaviour may be used against them, they may refuse to participate in the surveillance activities. Worse than this, they may participate and the results of the surveillance may lead to greater stigmatisation or crackdowns by law enforcement agencies. Other potential ethical issues and potential solutions are identified in Table 6.2 on the next page.

Although there are potential negative outcomes of identifying groups or behaviours in surveillance, there are also ethical issues in not including groups and individuals in prevention and treatment activities.

Working with high-risk groups, continued

Table 6.2. Potential ethical issues and solutions of being included in behavioural surveillance.

Potential ethical issues	Potential solution
Increases stigmatisation and discrimination of the group	• Ensure fully informed consent and absolute confidentiality • Reporting should be neutral • Reporting needs to be accompanied by public health communication about negative impact of stigma and discrimination on the epidemic
Loss of earnings	• Keep interview as short as possible • Remunerate lost earnings • Conduct interviews outside work times
Gatekeepers get angry at those who participate	• Involve and work with the gatekeepers, stress the benefits of surveillance
Gatekeepers force participation	• Involve and work with the gatekeepers, stress the benefits of surveillance
Illegal activities are highlighted, resulting in a police crackdown	• Involve and work with law enforcement agencies so they understand the purpose of surveillance and the damage that could result from them conducting repressive measures, such as scattering the high-risk groups or driving groups underground
Participants get no direct benefit from surveillance	• Report findings back to survey population • Explain the indirect benefits during the informed consent procedure • Do not foster false expectations

Discussing the table

Looking at Table 6.2, discuss the following questions:

1. What are some potential solutions when participants do not receive any direct benefits?

2. What are some potential ethical issues involving gatekeepers?

Working with adolescents

Different countries will have different laws and standards about when an adolescent can participate in research involving sexual behaviours, and when parental consent is required. Familiarise yourself with these laws in your country as part of your initial formative research efforts. Generally, surveillance tries to minimise the number of participants in the age range 15-18 and avoid including those under 15. If it is necessary to include children under the age of 15, special guidance on research with children should be sought.

Benefits to participants

In some countries, surveillance is accompanied by an intervention. Some would argue that surveillance is unethical in the absence of an intervention. However, usually incentives (the ethical issues of which are discussed above) are the only direct benefits of participating.

It is essential not to foster false expectations of direct benefits to participants. However, there are indirect benefits that should be stressed during the informed consent procedure and during community entry, including:

- Improving HIV prevention and care programmes
- Raising public awareness of and sympathy for burden of disease in the population
- Reducing stigma and effecting social change, especially around HIV infection
- Feedback of results to the community
- Incentives

Discussion

Looking at Table 6.3, discuss the following questions:

a. What are some possible incentives you could offer?

b. In what ways could you provide feedback of the results to the community?

Limits in ethical considerations

Fieldworkers may be asked for assistance from participants (for example, transporting a sick person to a health facility). There are limits to what is practical. Each country will need to decide this with reference to the local context. They should not make any promises they cannot keep.

Surveillance as research

Public health surveillance is not usually considered research and does not have the same requirements as research for documenting and reviewing ethical procedures. Although these ethical safeguards are not 'required,' they are still an essential part of surveillance. Your protocol should go through local ethics committees and institutional review boards (IRBs).

Summary

During behavioural surveillance, you should adhere to the universally accepted ethical principles of working with humans: respect for persons, beneficence and justice. A person can only make an informed decision to give his or her consent to participate when he or she has information about the procedures and purpose. Confidentiality protects subjects from adverse consequences that may arise from other people knowing their responses or about their participation. It is important to understand the potential ethical issues and some solutions of being included in behavioural surveillance.

Human subject reviews (IRBs) consider surveillance to be non-research when it involves routine case reporting. However, some IRBs and national authorities believe that periodic, cross-sectional surveys should be subject to standard review as research.

Unit 6 Exercises

Warm-up review

Take a few minutes now to look back at your answers for the warm-up questions at the beginning of the unit. Make any changes you want to make. We will discuss the questions and answers in a few minutes.

Small group discussion

Get into small groups by country, region or province to discuss these questions.

1. What are some of the potential social harms caused by behavioural surveillance in high-risk groups in your setting?

2. What ethical issues/difficulties have you/could you experience conducting surveillance in your country?

3. In a country with a concentrated epidemic, prevention programmes distributing clean needles have been hampered by police activity. Under local laws, the police may arrest people who are carrying needles, whether clean or used. In jail, heroin addicts suffering withdrawals become so desperate that they will inject with any needle they can obtain. As a result, one of the highest risks for HIV infection is incarceration (being jailed). Police activity is actually increasing HIV prevalence.

   a. Might this or a similar situation occur in your country?

   b. What changes would you recommend to reduce prevalence in your situation?

Apply what you've learned/ case study

Try this case study individually:

Design a consent form to be used with female sex workers.

Refer, if possible, to http://intranet.cdc.gov/od/ocso/osrs/hrpo/guides/consent/other.htm.
At that site, you will see sample consent forms in multiple languages.

Notes

Unit 7

Pre-Surveillance Activities

Overview

What this unit is about

In this unit, you will learn about the theoretical and practical knowledge required to plan behavioural surveillance activities. You will learn about the information that should be collected during the pre-surveillance assessment and the methods you can use to collect this information.

Warm-up questions

1. Which of the following is an example of a question you should answer in order to select appropriate geographical areas during the pre-surveillance assessment? Circle your answer.

    a. What are the regional differences in terms of HIV transmission and risk behaviour?
    b. Where are the interventions located?
    c. How much money and staff are available for surveillance?
    d. All of the above.

2. Match each pre-surveillance assessment method with its description.

_____	Assess what is currently known about the national epidemic or sub-epidemic.	a.	This step involves reviewing existing surveillance data, published and unpublished literature, and talking to people who are knowledgeable about the epidemic.
_____	Conduct pre-surveillance assessment using qualitative assessment methods and mapping.	b.	You can perform this step using existing information from general population surveys or collecting the information if none exists.
_____	Gather information on risk behaviours and HIV levels in the general populations.	c.	This step involves fieldwork to further identify hotspots and to gather information to define populations to be included in surveillance and guide fieldwork.

Warm-up questions, continued

3. Methods used in _____ research include large sample size, random samples and shorter interviews. Methods used in _____ research include unstructured questionnaire, lengthy interviews and fewer well-trained fieldworkers.

   a. qualitative
   b. quantitative

4. Suppose you want to determine in which areas of a city street-based sex occur. Which of the following is not an appropriate method to use during a pre-surveillance assessment?

   a. literature/data review
   b. in-depth qualitative interviews
   c. general population survey
   d. all are appropriate

Introduction

What you will learn

By the end of this unit, you should be able to:

- identify and understand the purpose of the pre-surveillance process
- understand and select the methodologies used in the pre-surveillance process.

Purpose of Pre-Surveillance Assessment

Identify populations and areas

The selection of surveillance populations should be based on a solid understanding of the epidemic dynamic in a country. The selection of geographical areas should reflect whether data are most meaningful if collected at a national or sub-national level. Several questions must be answered to gain this understanding.

Key questions to answer to select surveillance populations:

- Should surveillance concentrate on sub-populations, the general populations or both?
- Who are the most vulnerable populations (that is, those already infected with HIV or who have high-risk behaviours)?
- For each potential group, what is their potential contribution to the epidemic (how big are they, what links do they have to other populations, what types of risky behaviours do members engage in, with whom and with what frequency)?
- How will the populations contribute to understanding the epidemic?
- What interventions have been implemented or are planned among the populations?
- Is the population accessible for surveillance?

Key questions to answer in order to select geographical areas:

- What is already known about the epidemic in different regions of the country?
- What are the regional differences in terms of HIV transmission and risk behaviour?
- Where are the highest risk populations concentrated?
- What is the estimated size of the high-risk population in each geographic area?
- Are there cross border or internal transportation routes that could fuel transmission?

- Are the data needs at the national or sub-national level? At what level to we want to be able to generalise to?
- Can some areas serve as proxies for others?
- Where are the interventions located?
- Which locations have biological surveillance?
- How much money and staff are available for surveillance?

Determine feasibility

A successful surveillance system requires a feasible and effective data collection plan in order to enhance surveillance population participation, facilitate fieldwork and ensure appropriate sampling strategies are used. Several questions must be answered through the pre-surveillance assessment to ensure this.

Key questions to answer in order to ensure the participation of the surveillance populations and to facilitate fieldwork:

- Who are the gatekeepers of the populations?
- Whose permission is needed to conduct surveillance?
- Who can facilitate the survey process and make sure it runs smoothly and without disruption?
- In what language should the interview be conducted?
- What is the desired profile of the interviewers to ensure the most valid results?
- What are the possible locations where surveillance data could be collected where privacy will be ensured?
- What will be the best times of the day to find participants and to ensure that they will have time to complete the interviews?
- What practical problems could fieldworkers come across (for example, ensuring privacy, safety)?

Key questions to answer in order to ensure appropriate sampling strategies are used:

- Does a high proportion of the group gather at identifiable locations that can be listed and are they accessible through those locations?
- Do group members know each other? Are they part of a network?
- Can the same individuals be found at more then one location?
- Is it is possible to identify the members of the sub-population at the locations where they gather?

Define groups and eligibility criteria

One of the biggest sources of error in surveillance is a failure to track populations in a consistent manner over time. This sometimes happens because a population is not defined in sufficient detail. For example, there are many ways to define female sex workers. It is thus crucial to have a clear operational definition of what a female sex worker is and to have specific eligibility criteria of who can and cannot be included in surveillance.

Key questions to answer to define populations and set eligibility criteria:

- Are the risky behaviours in a population diverse (for example, for commercial sex workers how, when and where they operate)?
- Can the group be divided by differences in their behaviours and organisation?
- What characteristics can be used to identify group members?

Pre-Surveillance Assessment Methods

Three methods should be used for the pre-surveillance assessment:

1. Assess what is currently known about the national epidemic or sub-epidemic: This step involves reviewing existing surveillance data, published and unpublished literature, and talking to people who are knowledgeable about the epidemic to make a 'first cut' at identifying potential hotspots and at-risk or vulnerable populations.

2. Conduct pre-surveillance assessment using qualitative assessment methods and mapping: This step involves fieldwork to further identify and verify hot spots as well gathering information to clearly define populations to be included in surveillance and to guide surveillance fieldwork.

3. Gather information on risk behaviours and HIV levels in the general populations: This can be done using existing information from general population surveys or collecting the information if none exists. This information can be used to help validate whether the so called high-risk populations are really at higher risk than the population at large.

Methodological Details

Reviewing existing data

There are several sources of information that need to be reviewed. These include the peer-reviewed scientific literature, abstracts from regional and international AIDS conferences, *grey literature* (literature that is not published in easily accessible journals or databases, e.g., programme evaluations and governmental reports) and basic surveillance data.

Peer-reviewed literature can be located using internet-based search engines, such as Entrez PubMed from the U.S. National Library of Medicine. Particular care needs to be taken with specifying keywords and search terms.

- If terms are too general (for example, HIV), thousands of studies will be identified.
- If terms are too specific (for example, mentioning a specific city), nothing may be found.
- A good approach is to start with the name of your country and the keyword HIV.
- Additional search terms can be added, such as the name of the high-risk group.
- Once studies have been identified, accessed and reviewed, their bibliographies should be reviewed to identify other sources your search may not have identified.

Grey literature can also be located using the Internet. The UNAIDS website, http://www.unaids.org, is a resource for accessing this literature. UNAIDS compiles epidemiological fact sheets about each country involved in HIV/AIDS prevention programmes, as well as specific populations.

In addition, it is helpful to identify prevalence studies that report on the populations in question. A good example of this type of data can be found in the database at the United States Census website for HIV and AIDS surveillance. You will find various country profiles that examine the patterns and trends of the epidemic, as well as maps and tables that serve to summarise the statistics for each region in a streamlined format (http://www.census.gov/ipc/www/hivaidsn.html).

Governments, donors or non-governmental organisations produce monitoring and evaluation reports. These can sometimes be accessed through the internet. Often, you will need to contact governmental officials or representatives of the donor or non-governmental organisations to obtain copies.

Qualitative and quantitative research

Qualitative research focuses on the characteristics or quality of things, rather than the quantity. *Quantitative research* has powerful tools for the analysis of numbers, but researchers all know that the things counted are often qualitative categories or definitions.

For example, the number of AIDS cases a country has is dependent on the AIDS case definition. The case definition has changed several times and will probably continue to change. For the AIDS case definition, at least we can think of a way to get a standard definition: put the experts and the organisations in a room and don't let them out until they agree.

Along the spectrum of HIV disease, the point where AIDS is said to begin is completely definable by experts. However, when we talk about human behaviour, we don't have this luxury.

If you ask someone how many sexual partners they had in the last six months, the answer depends on what that person means by partner or means by sex. It also depends on whether they have any reluctance (or the opposite) to talk about these behaviours.

This is in addition to the normal problems of self-reported survey research, such as recall and reporting biases or the need to meaningfully translate questionnaires.

Luckily, these definitions and reluctances are not completely individualised. Groups tend to share these definitions and dispositions.

The questions we want to answer as part of our pre-surveillance activities do not require precise and generalisable (macro) quantitative measures of how many people have certain knowledge or perform a certain behaviour, which can be generalised on a population level. After all, the surveillance itself can do this.

Rather, it requires an in-depth (micro) knowledge of such things as:

- the performance of a behaviour (for example, sex work or injecting drugs)
- knowledge of a group's identity if it is so organised
- what people who are like them or in the group do
- an understanding of how and why they do it

Qualitative and quantitative research, continued

Micro knowledge can provide a description of the:

- material circumstances of the performance of the behaviour
- how that behaviour is organised
- what the participants and larger society think about the behaviour
- the people who perform it

Equally, this research can help explain how people think about the risk associated with the behaviour and how they (both individually and as a group) deny or change their behaviour to accommodate the risk.

Qualitative researchers often talk of the need to present an insider's view (an *emic* view) in order to understand what people mean when they reply to questions in an interview. Because people try to make sense of their lives and tie the way they live and how they think about it together, researchers also talk about the culture of a group.

Culture (as it is used here) is a kind of language that fuses the circumstances and rationales for behaviour into a logic that makes sense from that insider's point of view. From the outside, it may not appear to make any sense. For example, street children often deny the risk of AIDS from unprotected sex because they are so preoccupied with daily survival that an illness that might strike them in ten years seems irrelevant.

Difference between qualitative and quantitative methods

The differences between qualitative and quantitative methods outlined above translate into practical differences in methods, as shown in Table 7.1 below.

Table 7.1. Differences between qualitative and quantitative methods.

Quantitative	Qualitative
Large sample size	Small sample size
Random sample	Purposeful sample
Calculated sample size	Theoretical sample size
Structured questionnaire	Unstructured questionnaire
No deviation for interviewers	Interviewer exploration encouraged
Shorter interviews	Lengthy interviews
More fieldworkers, but they are less skilled	Fewer fieldworkers, but they are well-trained

Difference between qualitative and quantitative methods, continued

Imagine we want to determine if it is necessary to sub-divide commercial sex workers into sub-populations. In order to define the different types of commercial sex work, there would be little point interviewing people randomly selected from the population. Instead, we need to determine who could best provide us with this information.

Informants could include:

- sex workers
- NGOs who work with CSWs
- brothel/bar owners
- clients of sex workers and pimps

Once the types of informants are determined, interviewees should be selected for their competence as an informant. A good informant is:

- knowledgeable about the topic
- a person you can talk to easily
- someone who understands the information you need and is willing to give it to you

Saturation sampling

Unlike quantitative surveys, the sample size for qualitative research is not calculated prior to data collection.

A common way of determining sample size in qualitative research is known as *theoretical* or *saturation sampling*. This involves continuing to interview informants until no new information is learned.

We could, for example, ask informants to list all the different types of places they know where women sell sex. When the lists begin to be repetitive and informants are not providing new information, an adequate sample size has been reached. Among those involved in commercial sex work, the types of commercial sex workers can be considered common cultural knowledge and a complete list of types of workers can probably be determined from only a few good informants.

Once the different types of sex work have been identified, we need to determine if they need to be considered as separate surveillance groups. For this we need to understand how the different types of sex workers or different places where they operate affect risk and transmission. This is best explored using *open-ended semi-structured interviews*.

Open-ended semi-structured interviews

Open-ended questions have no answer choices from which respondents may select their response. Instead, respondents must create their own answers and state them in their own words. Using open-ended questions allows topics to be covered in more depth and can stimulate thoughtful responses, including suggestions from the respondent for the researchers, probing of people's memories and the clarification of positions.

Probing is often necessary to avoid incomplete, uninterruptable or irrelevant answers and interviews tend to be much longer and require more skill to conduct than interviews conducted as part of large sample social science surveys.

A semi-structured interview is one that uses a written list of open-ended questions and topics that need to be covered, although not necessarily in a particular order. Interviewers are encouraged to deviate from the guide when necessary to follow up new leads on the topic. These methods are in contrast to quantitative research, which tries to ensure that each informant is asked exactly the same questions in exactly the same way.

To continue our example, let us say we uncover two major kinds of sex workers in our discussions with expert informants and sex workers themselves: brothel-based and street-based. Before we classify one or both of them as a surveillance population we would want to know such things as:

- whether street-based sex workers work routinely or only occasionally
- whether they work the same streets or are constantly shifting in response to police or other pressure
- whether sex workers move back and forth between the two settings
- whether they have many clients or only a few
- other issues that might affect their risk profile and feasibility of selection as a surveillance population

Certainly epidemiological risk can only be determined in risk factor studies, but these preliminary studies identify the populations and methods to apply surveillance to.

Additional methods

Qualitative methods useful during pre-surveillance activities include:

- key informant/expert interviews
- focus groups and in-depth interviews
- observation

Ethical note: These activities are typically IRB exempt, but should be submitted.

Expert interviews

You may need to interview colleagues in the local offices of the Ministry of Health, those involved in behavioural and biological surveillance, local NGOs, clinics and donor agencies in your area that work with the HIV/AIDS epidemic. These are called expert interviews because they utilise respondents who are, by virtue of special training or their work with the population, experts and have equivalent status to the researchers. These can be confused with key informant interviews. Key informants are members of the group, and often become informal assistants to the researcher. Expert interviews utilise the semi-structured open-ended interviews described in the example above.

Focus groups and in-depth interviews

Both in-depth interviews and focus groups use semi-structured open-ended interviews. However, in-depth interviews are less structured, and permit the respondent to talk about a wide range of issues related to the topic and themselves. These interviews can be more like conversations, and the conversations permit rapport to be developed between the researcher and the respondent. In-depth interviews can be conducted over several sessions. In-depth interviews are conducted with members of the high-risk group rather than external experts such as government officials.

Focus groups are group interviews that are designed to initiate conversations within the group so that the researchers can listen to how these ideas are expressed. They are relatively less useful for a pre-surveillance assessment for behavioural surveillance, as their primary goal is to generate intervention ideas, to test materials and to determine social norms about things members of the group can talk about in front of other focus group participants.

Because the interviews are, in a sense, public, it may be difficult for a respondent to express private thoughts or talk about truly intimate behaviours. Similarly, the group interview can be manipulated by one or several influential members, and responses can be quite skewed.

Focus groups and in-depth interviews, continued

To encourage openness in focus groups, gather together people from similar backgrounds or experiences to discuss a specific topic. A group usually consists of 8-12 people and is guided by a facilitator and a record taker who takes notes. Because of the group dynamics involved, facilitators need to have special training and be quite skilful. This is another reason why expert and in-depth interviews are the most common interviews used for pre-surveillance assessment.

Direct observations

Direct observation is a somewhat more advanced technique that involves passively observing the high-risk population in question to determine where and when it congregates. For instance, street-based sex workers may be observed to determine in which areas of a city they work and the times that they tend to begin and end work. Observation may be associated with discussions or conversations with respondents being observed. When this is carried out over a sufficient time to be acknowledged and accepted by the group being observed, it is called participant observation.

PLACE protocol

The PLACE (Priorities for Local AIDS Control Efforts) protocol is a new rapid assessment tool used to identify and formalise information on high transmission areas. PLACE uses key informants to identify sites where people meet new sex partners, then interviews people at the site to characterise the site in each area and map sites. PLACE also interviews individuals socialising at the site to describe the characteristics of the people at the site. More information can be found at: www.cpc.unc.edu

Evaluating the Current State of Surveillance

The first step in improving a country's surveillance system is to evaluate the current system and identify what is working well and where gaps remain. As you assess your strengths and weaknesses, key issues to consider include:

- Are the goals and objectives of the surveillance system clearly stated?
- Are there standard protocols and mechanisms used to collect data?
- How effective are the various components of the surveillance system: data collection, questionnaires, etc.?
- Is the acquired information being used and adequately disseminated?
- See the MMWR Updated Guidelines for Evaluating Public Health Surveillance Systems for more information at http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm

Summary

It is important to understand the theoretical and practical knowledge required to plan behavioural surveillance activities. The first step in improving a country's surveillance system is to evaluate the current system and identify what is working well and where gaps remain. As you assess your strengths and weaknesses, consider the key issues. To collect information during the pre-surveillance assessment, there are several methods you can us

Unit 7 Exercises

Warm-up review

Take a few minutes now to look back at your answers for the warm-up questions at the beginning of the unit. Make any changes you want to make.

Small group discussion

Get into small groups by country, region or province to discuss these questions.

1. Read the sections about reviewing existing data, qualitative research and the PLACE methodology. Discuss how the methods might be used in your setting. If you have used these methods, share your experiences.

2. Identify three strengths and three weaknesses in your country's behavioural surveillance system. Keep in mind the surveillance steps as you identify their strengths and weaknesses. Key issues to consider are:
    - How effective are the various components of the surveillance system: data collection, questionnaires, etc.?
    - Is the acquired information being used and adequately disseminated?
    - Are the goals and objectives of the surveillance system clearly stated?
    - Are there standard protocols and mechanisms used to collect data?

    If you have never used these methods, discuss how they might be used in your country. Identify your high-risk groups first.

Apply what you've learned/ case study

Try this case study individually. We'll discuss the answers in class.

1. Make a list of all the things that need to be done before data collection begins.

2. What do you want to know before writing a protocol? How will you find out? Discuss fully.

Final Case Study

In this case study, answer the questions for each section before moving on and reading the section.

Part I

Global campaigns have focused on the fact that 50% of new HIV infections occur in people under the age of 24—an age at which people are discovering their sexuality and likely to be engaging in risk behaviour.

A country with a rapidly growing epidemic pooled their national surveillance data and analysed them by age. It was found that some 68% of all existing infections were among young people under 25.

A front-page article in the leading newspaper reports that behavioural surveillance among secondary school children had turned up shocking indicators: over half of the students had multiple sex partners and only 20% used condoms.[1]

As a ministry of health official:

- What specific programme(s) might you implement in response to these data?
- To what population(s) will you target your programs?
- What additional information do you need to make better programming decisions?

[1] Extract from the data-use module: Using data from HIV surveillance systems: Guidance on effective use of data from second generation surveillance systems: Elizabeth Pisani.

Part II

The Ministry of Health reacts rapidly, successfully lobbying for an allocation of US $4 million to develop life-skills curriculum focusing on helping young people to avoid unsafe sex.

The programme is implemented quickly, and two years later behavioural surveillance is repeated. The good news is that multiple partnerships among youth dropped from 50% to 25% and reported condom use doubles. However, the number of infections in young people reported by the HIV surveillance system has continued to grow.[2]

As a Ministry of Health official:

- How will you explain these data? What might account for them?
- What does this information tell you about your current HIV/AIDS programming? Would you make any changes?

[2] Extract from the data-use module: Using data from HIV surveillance systems: Guidance on effective use of data from second generation surveillance systems: Elizabeth Pisani

Part III

A parliamentary committee orders an inquiry. Public health officials go back to the behavioural surveillance data. It turns out that the indicators reported in the baseline year were accurate, in that half of all students who had had sex in the previous year had multiple partners and only 20% had used condoms. However, closer inspection of the data revealed that:

- Only 8% of all students had ever had sex at all, and only 4% had been sexually active in the past year
- 10% of all secondary school students had reported injecting drugs in the previous year and by the second round of behavioural surveillance that proportion had increased to 14%

These very high rates of injecting drug use had been overlooked because health officials, the press, and other agencies were looking for evidence of unprotected sex.

The ministry of health quickly ordered an assessment of injecting drug use and found that 85% of injectors were under the age of 25. Life-skills programs were subsequently redesigned to focus more on helping young people stay away from drugs. Harm reduction programmes were redesigned for those already injecting drugs, and injecting drug users were added as a group to the national surveillance system.[3]

What specific lessons can be drawn from this example?

[3] Extract from the data-use module: Using data from HIV surveillance systems: Guidance on effective use of data from second generation surveillance systems: Elizabeth Pisani

Notes

Summary

Surveillance is the systematic, regular and ongoing collection and use of data for public health action. Although they are often the beginning of a surveillance system, one-time cross-sectional surveys should not be considered surveillance.

Behavioural surveillance involves regular, repeated cross-sectional surveys collecting data that can be compared over time to HIV risk behaviours and other relevant issues. Biological surveillance also involves regular and repeated cross-sectional surveys, but collects biological samples that are tested for HIV and other related illnesses such as STIs and TB.

The uses of behavioural surveillance include the following:

- to provide an early warning of which groups and areas infection is likely to spread in and between
- to explain changes in HIV prevalence over time
- to provide information for developing prevention programmes
- to monitor and evaluate the impact of prevention programmes
- to reinforce the findings of biological surveillance
- to raise the awareness of HIV among policy-makers

Considerations when designing a behavioural surveillance system include:

- who to include in surveillance
- where to access the surveillance populations
- how to link biological and behavioural surveillance data
- how to ensure surveillance is appropriate for the context

The two most difficult issues defining behavioural surveillance indicators are defining the behaviours themselves and defining the time period for to the indicator should refer.

In observational studies, measurement error can come from:

- questionnaire faults
- interviewer error
- respondent error

Summary, continued

Data collection methods that can be used in behavioural surveillance include:

- face-to-face interviews
- self-administered questionnaire
- computer-assisted method

General population survey instruments are widely used. They should be considered to use solid sampling procedures and thorough statistical analysis. Most importantly, they should reflect a reliable source of behavioural data. Reliable general population survey instruments include:

- demographic and health surveys (MACRO)
- Multiple Indicator Cluster Survey (UNICEF)
- behavioural surveillance surveys (FHI)

The target population is the group that meets a survey's measurement objective (for example, all commercial sex workers in a city). The survey population is the target population modified to take into account practical considerations. (For example, all commercial sex workers in a city over the age of 15, excluding those who are based at home, as these cannot be accessed).

Issues for sampling include:

- consistent sampling is required across survey rounds
- general population surveys can rarely be used to access high-risk groups
- cluster sampling can be difficult when clusters are not stable
- members of high-risk groups may be difficult to identify and access
- cluster sampling is impossible if group members do not congregate

Behavioural surveillance sampling options include:

- conventional cluster sampling
- time-location sampling (TLS)
- respondent-driven sampling (RDS)

Summary, continued

Data analysis is summarising, presenting and interpreting data. Behavioural surveillance analysis can either be:

- cross-sectional - analysing data from one surveillance round
- trend - analysing data over several surveillance rounds

Informed consent means that you tell the person enough about the nature of the surveillance for them to make a proper (informed) decision about whether or not to take part.

Potential harms caused by behavioural surveillance in high-risk groups include:

- physical (public attack, abuse, loss of health-care services)
- legal (arrest, prosecution)
- social (disclosure to family, workplace, discrimination, loss of employment, isolation)

Potential benefits from behavioural surveillance include:

- improving HIV prevention and care programmes
- raising public awareness of burden of disease in the population, sympathy
- reducing stigma and effecting social change, especially around HIV infection
- feedback of results to the community
- incentives

The purpose of pre-surveillance assessment is to:

- identify appropriate surveillance populations and geographical areas
- determine feasibility of conducting surveillance in high-risk populations
- operationally define high-risk groups and set eligibility criteria for inclusion of population members in surveillance

Three methods should be used for the pre-surveillance assessment:

- Assess what is currently known about the national epidemic or sub-epidemic.
- Conduct pre-surveillance assessment using qualitative assessment methods and mapping.
- Gather information on risk behaviours and HIV levels in the general populations.

Notes

Appendix A, References and Further Reading

Attawell K. International longitudinal research on childhood poverty: practical guidelines and lessons learned from young lives. September 2003.

Buavirat A, Page-Shafer K, van Griensven GJP et al. Risk of prevalent HIV infection associated with incarceration among injecting drug users in Bangkok, Thailand: case-control study. *BMJ* 2003 Feb; 326:308.

FHI et al (2000). Behavioural surveillance surveys: guidelines for repeated behavioural surveys in populations at risk of HIV. Arlington, FHI.

FHI et al. The pre-surveillance assessment: Guidelines for planning serosurveillance of HIV, prevalence of sexually transmitted infections and the behavioural components of second generation surveillance of HIV, 2005

Gibney L, Saquib N, Metzger J. Behavioural risk factors for STD/HIV transmission in Bangladesh's trucking industry. *Soc Sci Med* 2003;56:1411-1424.

Kirkwood (1988). Essentials of medical statistics. Blackwell: Oxford.

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

Magnani et al (2001). Sampling strategies for monitoring HIV risk behaviours.

In Rehle et al. Evaluating programs for HIV/AIDS prevention and care in developing countries. FHI: Arlington.

Mehta SH, Gupta A, Sahay S, Godbole SV, Joshi SN, Reynolds SJ et al. High HIV prevalence among a high-risk subgroup of women attending sexually transmitted infection clinics in Pune, India. *J Acquir Immune Defic Syndr* 2006;4:75-80.

Mills et al (2004). Surveillance and modelling of HIV, STI, and risk behaviours in concentrated HIV epidemics. Sex Transm Infect 80(supp 2):57-62.

Mullany LC, Maung C, Beyrer C. HIV/AIDS knowledge, attitudes, and practices among Burmese migrant factory workers in Tak Province, Thailand. *AIDS Care* 2003;15:63-70.

Panda S, Kumar M, Lokabiraman S et al. Risk factors for HIV infection in injection drug users and evidence for onward transmission of HIV to their sexual partners in Chennai, India. *J Acquir Immune Defic Syn* 2005; 39:9-15.

Pisani E, Girault P, Gultom M, Sukartini N, Kumalawati J, Jazan S et al. HIV, syphilis infection, and sexual practices among transgender sex workers, male sex workers, and other men who have sex with men in Jakarta, Indonesia. *Sex Transm Infect* 2004;80:536-540.

Saengdidtha B, Lapparat G, Torugsa K, Suppadit W, Wakai S. Sexual behaviours and human immunodeficiency virus infection among Thai army conscripts between 1992 and 1998. *Mil Med* 2002;167:272-6.

Sarkar K, Bal B, Mukherjee R, Niyogi SK, Saha MK, Bhattacharya SK. Epidemiology of HIV infection among brothel-based sex workers in Kolkata, India. *J Health Popul Nutr* 2005; 23:231-5.

Semaan et al (2002). Street and network sampling in evaluation studies of HIV risk-reduction interventions. *AIDS Review* 4(2):213-23.

UNAIDS (2002). Initiating second-generation HIV surveillance systems: practical guidelines. WHO & UNAIDS: Geneva.

UNAIDS/WHO (2004). AIDS epidemic update: December 2004.

UNAIDS/WHO (2005). AIDS epidemic update: December 2005. Available at: http://www.unaids.org/epi/2005/

UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2004). Guidelines for the effective use of data from HIV surveillance systems.

UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2005). The pre-surveillance assessment: guidelines for planning serosurveillance of HIV, prevalence of sexually transmitted infections and the behavioural components of second generation surveillance of HIV, WHO/FHI.

Vajpayee M, Kanswal S, Seth P, Wig N. Spectrum of opportunistic infections and profile of CD4+ counts among AIDS patients in North India. *Infection* 2003 Oct;31:336-40.

WHO (2001*).* Guidelines for conducting HIV behavioural surveillance. WHO, SEA-AIDS-123: New Delhi.

Woodward & Chambers (1980). Guide to questionnaire construction and question writing. Canadian Public Health Association: Ontario.

Appendix B, Glossary and Acronyms

ACASI: Acronym for 'audio computerised assisted survey instruments.'

Accuracy: Refers to how well the sample reflects (nearest to the truth) the study population.

Acquired immunodeficiency syndrome (AIDS): See Advanced HIV infection.

Active infection: An infection that is currently producing symptoms (disease) or in which the organism that causes disease is reproducing.

Active surveillance: A system in which the organisation conducting surveillance initiates procedures to obtain reports. Example: making telephone calls or visits to health facilities to obtain information.

Adherence: The extent to which a patient takes his/her medication according to the prescribed schedule (also referred to as 'compliance').

Advanced HIV disease reporting: The systematic and standardized ongoing reporting of persons diagnosed with advanced HIV disease (clinical stage 3 or 4 and/or CD4 counts ≤350.

Advanced HIV infection: (*also* Advanced HIV disease) The late stage of HIV infection that includes development of one or more opportunistic illnesses (illnesses that occur because of low levels of CD4 lymphocytes, or immunodeficiency). Advanced HIV infection (disease) is the term now used for AIDS in updated WHO Guidelines.

Aetiologic case reporting: A surveillance system in which a laboratory test has confirmed the presence of the pathogen.

Aetiological: Refers to the causes of disease. Also known as 'aetiologic.'

Agent: A factor, such as a micro-organism, chemical substance, or form of radiation, whose presence is essential for the occurrence of a disease.

Aggregate case reporting: A single form summarises all of the patients who were diagnosed with the condition at certain sites in a given time period.

AIDS: Acronym for 'Acquired Immunodeficiency Syndrome.'

AIDS case reporting: The identification and reporting of persons meeting the AIDS case definition to permit public health authorities to track the disease over time. Also known as 'AIDS case surveillance.'

AIDS case surveillance: The identification and reporting of persons meeting the AIDS case definition to permit public health authorities to track the disease over time. Also known as 'AIDS case reporting.'

AIDS-defining illness: Any of a series of health conditions that are considered, in isolation, or in combination with others, to be indicative of the development of AIDS. These conditions result from low levels of CD4 lymphocytes which are destroyed by HIV.

AIDS Indicator Survey (AIS): A standardized tool to obtain indicators for effective monitoring of national HIV/AIDS programs. The protocols will help us provide, in a timely fashion and at a reasonable cost, the information required for meeting HIV/AIDS program reporting requirements.

Algorithm: Step-by-step procedure for decision-making; a recipe for achieving a specific goal.

Aliquot: A portion of a sample; for example, an aliquot of a 100 millilitre sample of blood might be a 5 millilitre portion of that sample.

Alliances: Partnerships created to assist with formative assessment. These partnerships differ based on the type of most-at-risk group being sampled, but usually include gatekeepers, governmental or non-governmental organisations, influential members of the target group, advocates, and physicians and others who provide health care to the target group.

Anonymous: Having no known name or identity. Removing all personally identifying information from a sample that will be tested for HIV, for example, in order to protect the patient's identity.

Anti-microbial resistance: The ability of an organism to avoid destruction or deactivation typically caused by drugs or chemicals designed to do so.

Antibiotic medicines: Drugs that kill or inhibit the growth of bacteria.

Antibodies: Molecules in the blood or secretory fluids that tag, destroy, or neutralise bacteria, viruses, or other harmful toxins.

Antimicrobial agents: An agent that kills or inhibits microbial growth. 'See Antibiotic medicines.'

Antiretroviral drugs: Drugs used to fight infections caused by retroviruses, such as Advanced HIV Disease.

Antiretroviral drug resistance: Resistance to one or more antiretroviral drugs. Antiretroviral drug resistance is one of the more common reasons for therapeutic failure in the treatment of HIV.

Antiretroviral therapy (ART): Treatment with drugs that inhibit the ability of HIV to multiply in the body.

Area map: A map used as a graph showing variables by geographic location.

Artefact: An inaccurate observation, effect or result caused by experimental error.

Asymptomatic: Without symptoms.

At-risk groups: Groups of people that are at increased risk for passing HIV on to others or for being infected by others.

B-lymphocytes: Also known as 'B-cells.' Blood cells of the immune system involved in the production of antibodies. In persons living with AIDS, the functional ability of both the B and the T lymphocytes is damaged, with the T lymphocytes being the principal site of infection by HIV.

Bacterial vaginosis: A chronic inflammation of the vagina caused by the bacterium *Gardnerella vaginalis*.

Bangui: The initial WHO AIDS surveillance case definition, developed to provide case definition of AIDS for use in countries where testing for HIV antibodies was not available.

Bar chart: A visual display of the size of the different categories of a variable. Each category or value of the variable is represented by a bar (or column). The Y-axis represents frequency. The X-axis represents different classes.

BED assay: A simple enzyme immunoassay (EIA) that can be used for detecting recent HIV-1 infection (within the last 160 days). It uses a branched peptide that includes sequences from HIV sub-types B, E and D, and allows detection of HIV-specific antibodies among various sub-types.

BED capture-EIA test: This test detects an antibody to a small HIV protein, gp41. It was first tested in HIV types B, E and D, hence its name BED.

Behavioural surveillance: Surveys of HIV-related behaviour that involve asking a sample of people about their risk behaviours, such as their sexual and drug-injecting behaviour.

Beneficence: To promote the interest of the patient or participant. To balance the benefits and risks to people involved in surveys. These risks include physical harm, such as violence and psychological harm, such as social stigmatisation.

Bias: A systematic error in the sample selection and the collection or interpretation of data.

Biological surveillance: Surveillance that involves regular and repeated cross-sectional surveys, but collects biological samples that are tested for HIV and other related illnesses, such as sexually transmitted diseases and tuberculosis.

Bivariate analysis: One of the main types of behavioural surveillance analysis that is performed to determine whether one variable is related to the distribution of another. For example, there might be an association between a respondent's age (the explanatory variable) and their use of condoms (the outcome variable). Variables are associated if the value of one tells you something about the value of another. Statistical tests in bivariate analysis determine whether any observed difference reflects a true difference, or may be due to chance.

Body fluids: Any fluid produced by the human body, such as blood, urine, saliva, sputum, tears, semen, mother's milk, or vaginal secretions. Fluids that commonly transmit HIV are blood, semen, pre-ejaculate, vaginal fluids, and breast milk.

Bridging populations: Persons in high-risk sub-populations who interact with people of lower risk in the general population, making it more likely that the HIV epidemic shifts from concentrated to generalised.

BSS: Acronym for ‘behavioural surveillance survey.’

***Candida albicans*:** The fungal causative agent of vulvovaginitis in women and inflammation of the penis and foreskin in men.

CAPI: Acronym for ‘computer-assisted personal interview.’

Capture-recapture: A technique used to estimate numbers of persons in a target population. Two or more lists containing individuals in common can establish the number of individuals missing from both, thereby estimating the total population of interest.

Carrier: A person or animal without apparent disease who harbours a specific infectious agent and is capable of transmitting the agent to others.

Case: An individual in the population or sample with a particular disease of interest.

Case-based reporting: each person diagnosed with the disease is reported separately, as opposed to aggregate case reporting in which data from patients with the disease are combined.

Case-control study: A type of observational analytic study. Enrolment into the study is based on presence (‘case’) or absence (‘control’) of disease. Characteristics such as previous exposure are then compared between cases and controls. The purpose of case

control studies is to identify factors that are associated with, or explain the occurrence of the specific disease or condition being studied.

Case definition: A set of standard criteria for deciding whether a person has a particular disease or health-related condition, by specifying clinical criteria and limitations on time, place and person.

Case fatality rate: The proportion of patients who become infected or develop a disease that dies as a result of that infection or disease.

Case reporting: A surveillance system in which persons who are identified as meeting the case definition are reported to public health authorities.

CASI: Acronym for 'computerised assisted survey instruments.'

Catchment population: A geographic area that is to be examined or surveyed. Can refer to the population served by a given clinic.

Categorical surveillance system: System that deals with reporting a single disease.

Categorical variable: Items that can be grouped into categories, such as marital status or occupation.

Cause of disease: A factor (characteristic, behaviour, etc.) that directly influences the occurrence of disease. A reduction of the factor in the population should lead to a reduction in the occurrence of disease.

CD4 count: A measure of the number of CD4 cells in a millilitre (mL) of blood. The CD4 count is one of the most useful indicators of the health of the immune system and a marker for the progression of HIV/AIDS.

CD4 receptors: Markers found on the surface of some body cells, including T-cells. These receptors are targets of HIV, and thus CD4+ cells are attacked by the virus.

Census sampling: Every unit, or case, is measured for the entire population. A de facto census allocates persons according to their location at the time of enumeration. A de jure census assigns persons according to their usual place of residence at the time of enumeration (Last).

Centers for Disease Control and Prevention (CDC): The US Department for Health and Human Services agency with the mission to promote health and quality of life by preventing and controlling disease, injury, and disability.

Chain referral sample: Any sampling method wherein participants refer other potential participants for inclusion in the sample. There are several types of chain referral sampling

methods, most of which are non-probability samples. Examples of chain referrals include RDS, network sampling, random walk and snowball sampling.

Chancroid: An acute, sexually transmitted, infectious disease of the genitalia caused by the bacteria *Haemophilus ducreyi*. The infection produces a genital ulcer that may facilitate the transmission of HIV.

Characteristic: A definable or measurable feature of a process, product, or variable.

Chlamydia trachomatis: The most common sexually transmitted bacterial species of the genus Chlamydia that infects the reproductive system. Chlamydia infection causes infection of the cervix of women and the urethra of men and is frequently asymptomatic. If left untreated, it can cause sterility in women.

Clinic-based surveys: Surveys that use samples that have been selected in clinical facilities, such as STI or drug treatment clinics. The most common type of the clinic-based surveys that are done using biological markers, such as HIV infection, is clinic-based sentinel sero-surveillance.

Cluster: Any aggregate of the population of interest (for example, departments, villages, health facilities).

Cluster sampling: The population of interest is broken into groups or clusters and a sample of clusters is randomly selected (Levy & Lemeshow).

Clustered bar chart: A bar chart in which the columns are presented as clusters of sub-groups. Also known as 'stacked bar chart.'

Code: A unique identification for a specimen. It may or may not be linked to any personal identifying information.

Cohort analysis: Analysis that involves following groups of subjects over time.

Cohort studies: Cohort studies follow a group of initially uninfected people over time, and test them repeatedly. Cohort studies follow a well-defined group of people who have had a common experience or exposure, who are then followed up for the incidence of new diseases or events, as in a cohort or prospective study tested repeatedly over a long period of time.

Community advisory board: Members of the community who offer input into study design and local procedures. CAB members include community activists and/or professionals associated with HIV/AIDS prevention and services delivery. Some CAB members are trial participants.

Community-based surveys: Surveys that use samples that have been selected from non-clinical settings. They often include most-at-risk populations, such as sex workers or

truck drivers, who are not included in clinic-based surveys. As with clinic-based surveys, the most common type of community-based survey is called 'repeated cross-sectional community-based sentinel sero-surveillance.'

Community sites: Locations in the community, such as households or brothels.

Completeness of data elements: The extent to which the information requested in the case report form is provided.

Completeness of reporting: One of several attributes of a surveillance system. The term refers to the proportion of cases that were reported. Completeness of reporting is also referred to as the sensitivity of the surveillance system and is determined by using an alternative (and thorough) method of identifying cases of the disease and then dividing the number of cases reported by the total number of cases identified. Completeness is often reported as a percentage.

Compulsory testing: Testing that is required of all individuals in a population to be surveyed. For example, requiring HIV tests to be done on all members of a prison population.

Concentrated HIV epidemic: The epidemic state in which HIV has spread to a high level in a defined subpopulation but is not well established in the general population. HIV prevalence is consistently >5% in at least one defined subpopulation and is <1% in pregnant women in urban areas.

Confidence interval: The compound interval with a given probability, for example, 95% that the true value of a variable such as mean, proportion, or rate is contained within the limits. Also known as 'confidence limits.'

Confidence limits: See 'confidence interval.'

Confidentiality: Protecting information that concerns a study participant or patient from release to those who do not need to have the information.

Consecutive sampling: This sampling method consists of sampling every patient who meets the inclusion criteria until the required sample size is obtained or the survey period is over. While this method is not strictly a probability sample, it is easier to use and offers less occasion for sampling bias.

Contact: Exposure to a source of an infection, or a person so exposed.

Contagious: The characteristic of an organism or person that renders it capable of being transmitted from one person to another by contact or close proximity.

Continuous variable: Items that occur in a numerical order, such as height or age.

Convenience sampling: The selection of entities from a population based on accessibility and availability. Available participants may be people on the street, patients in a hospital or employees in an agency. This type of sampling does not generally represent the population of interest and is best used in the exploratory stage of research.

Core data elements: Information about a patient that must be collected during a survey.

Cotrimoxazole preventative therapy (CPT): Administering cotrimoxazole prophylaxis to prevent opportunistic infections among HIV- infected patients.

Cotrimoxazole prophylaxis: A combination of two anti-infection drugs, sulfamethoxazole and trimethoprim, used to prevent opportunistic infections in patients with HIV.

Coupon: Used in RDS studies to provide incentives to participants. Coupons in RDS can be used both to track participation for reimbursements and to link the recruiters to the recruits. Other methods may use coupons to encourage participation, much like the advertisements placed in popular clubs or bars. Some coupons may have two parts that can be easily separated. One part of the coupon serves as the referral coupon, which the recruiter uses to recruit a peer into the study. The other part of the coupon serves as the payment coupon. It is kept by the recruiter and he or she will use it to claim an incentive for having recruited a peer into the study. Both parts of the coupon have the unique identification number of the recruitee printed on them. The dual system eliminates the need to collect names for incentive collection.

Coupon rejecters: People who are offered a coupon by a recruiter, but decline to take it.

Cross-sectional survey: A survey that is conducted over a given period of time, such as during a single year, rather than over an extended period of time.

Cruising area: Cruising areas are public space, such as parks, public restrooms, bath houses, dance clubs and railway stations where MSM meet, congregate and arrange and/or engage in sexual activity.

Cryolabel: Labels designed to adhere during freezer storage.

Cryovial: A vial that is designed to be stored in a freezer.

CSW: Acronym for ‘commercial sex worker.’

DALYs: See ‘disability-adjusted life years.’

Database: A computer programme that stores the variables for each patient in the survey sample or surveillance system.

Data dictionary: Electronic files that describe the basic organisation of a project or database. They contain all of the rules that guide data entry.

Data entry: The process of entering paper records into a computer database

Data entry screens: The forms on the computer screen into which a data entry clerk enters the data.

Data synthesis: See 'triangulation.'

Definitive diagnosis: A diagnosis based on laboratory or other tests specifically designed for diagnosis and considered authoritative.

Demographic Health Survey: National household surveys that provide data for a wide range of monitoring and impact evaluation on topics including HIV prevalence and attitudes and beliefs about HIV/AIDS.

Demographic information: The 'person' characteristics of epidemiology (usually collected with "place" and "time") – age, sex, race and occupation – used to characterise the populations at risk.

Denominator: The population (or population experience, as in person-years, etc.) at risk in the calculation of a proportion or rate. The denominator is the lower portion of a fraction used to calculate a rate or ratio.

Dependent variable: In a statistical analysis, the outcome variable(s) or the variable(s) whose values are a function of other variable(s).

Descriptive statistics: Used to describe the basic features of the data, they provide simple summaries about the sample and the measures.

DHS: Acronym for 'demographic and health surveys.'

Dichotomous variable: A special type of nominal variable that has only two categories, such as male/female.

Differential recruitment: Recruiters successfully bring recruits in at different rates.

Direct transmission: The immediate transfer of an agent from a reservoir to a susceptible host by direct contact or droplet spread.

Disaggregated data: Data which is divided up according to different variables, to provide a more detailed analysis.

Disability-adjusted life years (DALYs): A measure of burden of disease in a population obtained by combining 'years of life lost' and 'years lived with disability.'

Disease burden: The size of a health problem in an area, as measured by cost, mortality, morbidity or other indicators.

Disease registry: The file of data that contains reported diseases.

Disease reporting: The process by which notifiable diseases are reported to the health authority.

Disinhibition: Poor decision-making when considering risk-taking behaviours.

Distribution: The frequency and pattern of health-related characteristics and events in a population. In statistics, the observed or theoretical frequency of values of a variable.

Double-entered: Entered twice, to avoid mistakes by identifying and correcting discrepancies.

Double Y-scale: On a graph, two Y-axes, one on the vertical left for data with large values and one on the vertical right for data with smaller values.

Dysuria: Painful, frequent or difficult urination.

EIA: See 'enzyme-linked immunoassay.'

ELISA: See 'enzyme-linked immunosorbent assay.'

Emic: Refers to accounts, descriptions, and analyses expressed in terms of the concepts and categories regarded as meaningful and appropriate by the members of the population of interest.

Endemic disease: The constant presence of a disease or infectious agent within a given geographic area or population group; may also refer to the usual prevalence of a given disease within such area or group.

Enumeration units: The sampling units from the final stage of a multistage sampling design. See 'Listing units.'

Enzyme immunoassay (EIA): A type of test that identifies antibodies to an organism such as HIV. EIAs rely on a primary antigen-antibody interaction and can use whole viral lysate of HIV or one or more antigens from the virus.

Enzyme-linked immunosorbent assay (ELISA): A type of enzyme immunoassay (EIA) to determine the presence of antibodies to an infectious agent such as HIV in the blood or oral fluids.

Epidemic: The occurrence of a disease (or other health-related event) at a greater than expected level of increase to a baseline. For example, the high prevalence of HIV found in many parts of the world today, including sub-Saharan Africa, Latin America and South and Southeast Asia.

Epidemic state: The prevalence the epidemic has reached in a country or region. Can be low-level, concentrated, or generalized within a sub-population or within the general population.

Epidemiology: The study of the distribution and determinants of the frequency of health-related states or events in specified populations, and the application of this study to the control of health problems.

Epi Info™: Freely distributed epidemiological software available on the CDC website (www.cdc.gov/epiinfo).

Equilibrium: In RDS, the point in the recruitment process where a variable is not expected to change by more than 2% with each successive wave.

Ethnographic assessments: Ethnographic assessments are written analyses of the cultural practices, beliefs and behaviours of a particular culture, network or sub-group.

Ethnographic mapping: Collecting information on the geographic location, temporal movement of and interactions among members of the study population.

Etic: Refers to accounts, descriptions and analyses expressed in terms of the concepts and categories regarded as meaningful and appropriate by the community of scientific observers.

Exclusion criteria: Characteristics of patients who should be excluded from the sample, but who would otherwise be eligible.

Experimental study: A study in which the investigator specifies the exposure category for each individual (clinical trial) or community (community trial), then follows the individuals or community to detect the effects of the exposure.

External validity: The ability to make inferences from the study sample to the population of interest.

Factor: An intrinsic factor (age, race, sex, behaviours, etc.) which influences an individual's exposure, susceptibility, or response to a causative agent

False negatives: Test results that are negative when the patient actually has the disease that is being tested for.

False positives: Test results that are positive when the patient does not actually have the disease that is being tested for.

Female sex workers: Females who engage in sex work, or the exchange of sex for money, which includes many practises and occurs in a variety of settings. These may include 'direct' or 'formal' sex workers, who are sometimes included in registries and often found in brothels, and 'indirect' or 'casual' sex workers, who do not engage in sex work full time and are unlikely to be included in registries.

Filter paper: Porous paper on which samples can be placed.

Focus groups: A group setting in which people are asked by a facilitator about their views about a topic. Participants are free to talk with other group members as well as the facilitator. Focus groups allow interviewers to study people in a more natural setting than they can in a one-to-one interview.

Formative assessment (or research): Research conducted before the study begins. Researchers use qualitative methods, such as focus groups, in-depth interviews, mapping or observations of the target population and the individuals who work with them to ensure that the research team sufficiently understands the community.

GAP: Acronym for the CDC's 'Global AIDS Program.'

Gatekeepers: Persons who can provide access to a high-risk population. Examples are a brothel owner who can provide access to female sex workers, or a prison warden who can provide access to prisoners.

General population surveillance: Surveillance that measures HIV risk behaviours in a sample of people selected to represent the people living in a region or nation. The surveillance can be restricted to certain ages (for example, young people aged 15-24) or genders.

Generalisability: The results from the sample are the same as the results we would have obtained had we tested every person in the study population (that is, the results from the sample are generalisable to the study population).

Generalised HIV epidemic: The epidemic state in which HIV is firmly established in the general population. HIV prevalence is consistently >1% in pregnant women.

Genital discharge syndrome: This syndrome includes infections due to *N. gonorrhoea*, and *C. trachomatis*.

Genital ulcer syndrome: Genital lesions due to *T. pallidum*, *H. ducreyi*, HSV, *C. trachomatis* or *C. granulomatis*.

Geographical Information System (GIS): System of hardware, software.

Gigolo: Male sex workers who identify as straight. They tend to have foreign clients and engage in male-male sexual activity.

Glycoprotein (HIV): Proteins on the surface of the HIV virus that bind to CD4 receptors on target cells.and procedures designed for integrated storing, management, manipulation, analysing, modelling and display of spatially referenced data for solving planningand management problems.

Gonorrhoea: An infection caused by *Neisseria gonorrhoeae* bacteria. Although gonorrhoea is considered primarily a sexually transmitted infection, it can also be transmitted to newborns during the birth process.

Gram-negative: Bacteria that do not absorb the stain during the process of Gram staining.

Gram-positive: Bacteria that do absorb the stain during the process of Gram staining.

Gram stain: A laboratory method of staining microscopic slides of organisms in order to identify and classify the various types of bacteria. Bacteria are classified as either Gram-negative (does not absorb the stain) or Gram-positive (absorbs the stain).

Graph: A diagram that shows a series of one or more points, lines, line segments, curves or areas, representing variations of a variable in comparison with variations of one or more other variables.

Grey literature: Material that is not published in easily accessible journals or databases. Besides programme evaluations, government surveillance reports and programme planning documents mentioned earlier, it includes the abstracts of research presented at conferences, and unpublished theses and dissertations.

***Haemophilus ducreyi*:** The causative agent of chancroid. See ‘chancroid.’

Health indicator: A measure that reflects, or indicates, the state of health of persons in a defined population; for example, the infant mortality rate.

Health information system: A combination of health statistics from various sources, used to derive information about health status, healthcare, provision and use of services, and impact on health.

Health-seeking behaviour: The actions individuals or populations take to care for their health, for example, attending a clinic or district hospital when they feel ill.

Hard-to-reach populations (HTRP): Groups of people linked by behaviours, socioeconomic situations or societal structures, who for various reasons (e.g. law, stigma) refrain from involvement in the legal economy and other aspects of the majority social

institutions. Includes but is not limited to: IDUs, MSM, CSW and undocumented migrants.

Hepatitis B virus (HBV): The causative agent of hepatitis B. The virus is transmitted by sexual contact, the use of contaminated needles and instruments and by contaminated serum in blood transfusion. The infection may be severe and result in prolonged illness, destruction of liver cells, cirrhosis or death.

Hepatitis C virus (HCV): The causative agent of hepatitis C. This virus is transmitted largely by the use of contaminated needles and instruments and by blood transfusions. The disease progresses to chronic hepatitis in up to 50% of the patients acutely infected.

Herpes simplex virus 1 (HSV-1): A virus that causes cold sores or fever blisters on the mouth or around the eyes, and can be transmitted to the genital region.

Herpes simplex virus 2 (HSV-2): A virus causing painful sores of the anus or genitals. While this is a sexually transmitted infection, it may be transmitted to a newborn child during birth from an infected mother.

Herpes viruses: A group of viruses that includes herpes simplex type 1 (HSV-1), herpes simplex type 2 (HSV-2), cytomegalovirus (CMV), Epstein-Barr virus (EBV), varicella zoster virus (VZV), human herpes virus type 6 (HHV-6), and HHV-8, a herpes virus associated with Kaposi's sarcoma.

Highly active antiretroviral therapy (HAART): The use of at least three ARV drugs in combination to suppress viral replication and progression of HIV disease by reducing the viral load to undetectable levels.

High-risk behaviours: Behaviours that increase the risk that a person will contract a disease.

High-risk group: A group in the community with an elevated risk of disease, often because group members engage in some form of risky behaviour.

High-risk group surveillance: Surveillance that measures HIV risk behaviours in groups whose behaviours, occupations or lifestyles could expose them to higher risk of acquiring and transmitting HIV than the rest of the population. These groups are often important in establishing, accelerating or sustaining the HIV epidemic.

High-risk heterosexuals (HRH): Includes but is not limited to: mobile populations, uniformed personnel and sex partners of other MARPs.

Histogram: A graph that represents a frequency distribution by means of rectangles whose widths represent class intervals and whose areas represent corresponding frequencies.

HIV: See 'Human Immunodeficiency Virus.'

HIV-1: A type of HIV with slight genetic variations from HIV-2. More easily transmitted than HIV-2.

HIV-2: A type of HIV with slight genetic variations from HIV-1. Less easily transmitted than HIV-1.

HIV case reporting: the systematic, standardized, ongoing collection of reports of persons diagnosed with HIV infection (clinical stages 1-4) and/or advanced HIV disease (clinical stages 3 and 4).

HIV clinical stages: In these modules, a classification by WHO of HIV disease on the basis of clinical manifestations that can be recognized and treated by clinicians in diverse settings, including resource-constrained settings. In order of severity, starting with the lowest, the stages are:

Stage 1: Often asymptomatic or with swollen glands
Stage 2: Symptoms, including moderate weight loss and respiratory infections
Stage 3: More severe symptoms, including extreme weight loss and severe bacterial infections. Called advanced HIV disease.
Stage 4: End-stage HIV infection (AIDS), with manifestations such as wasting syndrome, tuberculosis, lymphoma. Called advanced HIV disease.

HIV-negative: Showing no evidence of infection with HIV (for example, absence of antibodies against HIV) in a blood or tissue test.

HIV-positive: Showing indications of infection with HIV (for example, presence of antibodies against HIV) based on a test of blood or tissue.

HIV sub-types: Distinct lineages of HIV that contain genetic differences.

HIV viral suppression: Lowering the level of HIV RNA in plasma, below the threshold of detection.

Homophily: In RDS, a measure of the tendency of people to connect to other people like themselves.

HSV-2: see herpes simplex virus 2.

Human immunodeficiency virus (HIV): A retrovirus that causes AIDS by infecting T-cells of the immune system.

Human papilloma virus (HPV): A causative agent of genital warts.

IDSR: See ‘Integrated disease surveillance and response.’

IDU: Acronym for 'injection (injecting or intravenous) drug user.'

Immune response: The activity of the immune system against foreign substances such as infectious agents including bacteria and viruses.

Immune system: The body's complicated natural defence against disruption caused by invading foreign agents (for example, microbes or viruses).

Immunodeficient: A situation in which a patient's health is compromised because his/her immune system is insufficient to ward off infections, thus making the person susceptible to certain diseases that they would not ordinarily develop.

Immunology: The science of the system of the body that fights infections.

Impact evaluation: An evaluation of a programme that determines what the impact of the programme is, as opposed to 'process evaluation.'

Impact indicators: A standardised set of indicators developed by UNAIDS to help monitor HIV prevalence in particular populations.

Incentive: A reward or reimbursement given to participants in a study. In RDS surveys, there are typically two levels of incentive: primary incentive and secondary incentive. A participant receives the primary incentive for enrolling in the study and completing an interview. The same participant receives secondary incentive(s) for recruiting his or her peers into the study. Incentives are not absolutely necessary in every situation and should be determined during formative research.

Incidence: A measure of the frequency with which an event, such as a new case of illness, occurs in a population over a period of time. The denominator is the population at risk; the numerator is the number of new cases occurring during a given time period.

Inclusion criteria: Characteristics required in study participants, in order to be considered for the sample.

Incubation period: A period of sub-clinical or unapparent pathologic changes following exposure, ending with the development of the infection.

Independent variable: An exposure, risk factor, or other characteristic being observed or measured that is hypothesised to influence the outcome (that is, the dependent variable).

Indicators: Specific data that are gathered to measure how well a prevention or treatment programme is doing as well as define an aspect of behaviour that is key to the spread of HIV. Indicators provide a way to track changes in behaviours over time and provide a way to compare levels of risk behaviours between different population groups.

Indicator mutations: Genotypic mutations that best predict resistance to a specific antiretroviral agent.

Indirect transmission: The transmission of an agent carried from a reservoir to a susceptible host by suspended air particles or by animate (vector) or inanimate (vehicle) intermediaries.

Infectiousness: The ability of an organism to cause infection.

Infectivity: The proportion of persons exposed to a causative agent who become infected by an infectious disease.

Information bias: Error that results from people who have a disease being misclassified as not having the disease.

Informed consent: The permission granted by a patient or a participant in a research study after he or she has received comprehensive information about a research study or medical procedure. Informed consent protects the person's freedom of choice and respects his or her autonomy with regard to decisions affecting his or her body and health.

In-group affiliation: In RDS, what homophily measures (group similarity based on ethnicity, age, socio-economic status and so forth).

Injection drug users (IDUs): Also called 'intravenous drug users,' they are persons who use or have used needles or syringes to inject drugs. Injection drug use is considered a high-risk behaviour.

Institutional review board (IRB): The committee designated to approve, monitor, and review biomedical and behavioral research involving humans with the aim of protecting the rights and welfare of research participants. Also known as ethics committee.

Institutional sampling: Individuals in an institution, such as prison, are sampled.

Integrated disease surveillance (IDS): An approach to surveillance in which communicable diseases are prioritised. Surveillance for all of the high-priority diseases is conducted in an integrated manner and is initiated at the district level. These diseases have a high potential for epidemic spread and can be controlled through public health measures.

Internal validity: The absence of substantial differences between groups at baseline; the absence of substantial difference of attrition rates between groups at follow-up.

Internally displaced persons (IDPs): IDPs are persons who have left their homes due to civil unrest or natural disasters, but have stayed in their homeland and have not sought sanctuary in another country.

Interval width: The range of certainty as to the true value of the calculated outcome value. For example, in the case of a 95% confidence interval, there is 95% certainty that the true outcome lies between the upper and lower bound of the interval. Statistically, this interval is equal to two standard deviations on either side of the calculated outcome value.

Interviewer error: Problems stemming from the actions and behaviours of the person doing the interview.

Intradermally: Injected into the layers of the skin.

Intramuscularly: Injected into a muscle.

Intravenously: Injected into a vein.

Involuntary migrants: Involuntary migrants include persons who have migrated away or have been displaced from their home countries due to an established or well-founded fear of persecution, or have been moved as a result of deception or coercion.

Isolates: A population of bacteria or other cells that has been isolated and cultured.

Isoniazid prophylaxis: Giving isoniazid to individuals with latent Mycobacterium tuberculosis infection, in order to prevent the progression to active disease. Prophylaxis with isoniazid significantly reduces the incidence of tuberculosis in adults with HIV and a positive tuberculin skin test result.

Key informants: Members of the target group, who can often become informal assistants.

Kick-off meeting: A meeting you host for community members who may in turn become seeds for the RDS survey. The purpose of the meeting is to educate seeds on study goals and process, inform seeds of their importance to the success of the study and encourage the seeds to be enthusiastic.

***Klebsiella granulomatis*:** The bacterial causative agent of granuloma inguinale or donovanosis.

Laboratory-initiated reporting: A surveillance system in which the reports of cases come from clinical laboratories.

Laryngeal TB: Tuberculosis involving the larynx, producing ulceration of the vocal cords and elsewhere on the mucosa, and commonly attended by hoarseness, cough, pain on swallowing, and hemoptysis.

Latent period: A period of unapparent infection following exposure to a pathogen, ending with the onset of symptoms of chronic disease.
Lessons learned: Information from actual studies that will help you make decisions when planning your study.

Linked anonymous HIV testing: In linked anonymous testing, a person agrees to have an HIV test, but the specimen is labelled with a code without a name or identifiers that could reveal the person's identity. This method is voluntary and requires obtaining informed consent and making the test results available (with appropriate counseling) to the person tested.

Linked confidential HIV testing: In linked confidential testing, a person agrees to have an HIV test with the assurance that the test result will be kept confidential and only selected health-care providers may be informed. This method is voluntary and requires obtaining informed consent and discussing the test results with the person. Linked confidential testing also allows for the collection of more detailed demographic and risk-behaviour information.

Linking: Refers to whether a tested individual's names or identifying information is associated with his or her HIV test results.

Listing units: The sampling units from the final stage of a multistage sampling design. See enumeration units.

Log scale: In a graph, when the data covers a large range of values, they are presented on a logarithmic scale. This type of scale reduces data to a smaller range so that it is easier to work with.

Longitudinal surveillance: Surveillance over time during which patients' status can be updated. *Longitudinal databases* allow the update of patients records over time with, for example, start dates for care, disease progression, new information.

Low-level HIV epidemic: The epidemic state in which HIV has never spread to significant levels in any sub-population, although HIV infection may have existed for many years. HIV prevalence has not consistently exceeded 5% in any defined sub-population. This state suggests that networks of risk are rather diffuse or that the virus has only been recently introduced.

Lymphocytes: A type of white blood cell that is involved with fighting infections in the body. The T lymphocyte is the cell that HIV infects and destroys.

Macrophage cells: Tissue cell derived from monocytes that protect the body against infections.

Male sex workers: Males who engage in sex work, or the exchange of sex for money, which includes many practises and occurs in a variety of settings.

Mandatory testing: Testing that is required of a patient if he or she is to obtain certain services; for example, mandatory HIV testing of individuals who request marriage certificates.

Margin of error: An estimation of the extent to which a survey's reported percentages would vary if the same survey were taken multiple times.

Markov process: A mathematical theory that provides a probabilistic description of the state of a system at any future time. The Markov process is especially relevant to RDS because of the nature of the recruitment process, whereby a chain of peers recruiting peers is monitored through a coupon mechanism.

Marriage pressure: Family pressure on sons to marry to provide stability for parents and the continuation of the family name as well as to avoid the stigma of a person being MSM.

MARP: Acronym for most-at-risk population, a group within the community with an elevated risk of disease, often because group members engage in some form of high-risk behaviour.

Masking: Describes the behaviour of reclusive respondents, people who do not want to be found.

Mean: The measure of central location commonly called the average. It is calculated by adding together all the individual values in a group of measurements and dividing by the number of values in the group.

Men who have sex with men (MSM): Men who have sex with men (MSM) are one of the highest risk groups in the Americas, Asia, Europe and Oceania. For the purposes of this manual, we also consider male sex workers, transvestites and transgendered persons (*hijra*) in the MSM category.

Microbe: A micro-organism, such as a bacteria or virus.

Microbicide: A chemical or other agent that destroys microbes.

MICS: See 'Multiple Indicator Cluster Survey.'

Migrants: see 'mobile populations'

Mobile populations: Refers collectively to groups of people who move from one place to another (migrants). They may move temporarily, seasonally, or permanently and for either voluntary or involuntary reasons.

Monitoring: Evaluating a programme's performance over time.

Monitoring and Evaluation (M&E): Collecting and analysing accurate and reliable information that can be used to improve programme performance and planning.

Monocyte: A type of white blood cell.

Morbidity: Any departure, subjective or objective, from a state of physiological or psychological well-being.

Mortality rate: A measure of the frequency of occurrence of death in a defined population during a specified interval of time.

Mortality rate, infant: A ratio expressing the number of deaths among children under one year of age reported during a given time period divided by the number of births reported during the same time period.

MSC: See 'multi-stage cluster sampling.'

MSM: Acronym for 'men who have sex with men.'

MSW: Acronym for 'male sex worker.'

MTCT: Acronym for 'mother-to-child transmission.' See 'perinatal transmission.'

Multi-stage cluster sampling (MSC): Two- or more- stage sampling. Final units from selected clusters may be randomly selected.

- Simple two-stage cluster sampling
- Probability proportional to size sampling (PPS) is used when all clusters do not have the equal probability of being selected in the sample. PPS is a class of unequal probability sampling in which the probability of a unit being sampled is proportional to the level of some known variable (Levy & Lemeshow).

Multivariate analysis: One of the main types of analysis conducted in behavioural surveillance that is performed to look at the influence of at least two variables on another variable. since relationships between variables are often complex and interwoven. Multivariate techniques can pinpoint the individual effects of several explanatory variables on an outcome variable, which may be related to each other.

Natural history of disease: The temporal course of disease.

Needs assessment: A systematic examination of the type, depth and scope of a problem.

Negative controls: Specimens known to be negative and used to ensure that a laboratory reagent is working properly prior to testing specimens from patients.

Negative predictive value: In HIV testing, the probability that a person with a negative test result is not infected. Also known as 'predictive value negative.'
***Neisseria gonnorrhoeae*:** The causative agent of gonorrhoea.

Network: This sampling method may be used for groups whose members are socially linked. Ego-centred network sampling is based on random, representative or any other form of quota sampling (Schensul). Full relational network sampling begins with identification of individuals (see 'seeds') who act as entry points to the network.

NGO: Acronym for 'non-governmental organisation.'

Nominal variable: Variables that represent discrete categories without a natural order, such as marital status.

Non-probability sampling: The sampling units are selected through a non-randomised process; therefore, the probability of selecting any sampling unit is not known.

Non-random mixing: The tendency of people to associate preferentially with others who are like themselves.

Non-vesicular genital ulcer disease: An STI syndrome characterised by ulcers and the absence of vesicles.

Notifiable disease: A disease for which law or regulation requires reporting to the health authority.

Numerator: The upper portion of a fraction. In a rate, the numerator is usually the number of people infected.

Operational definitions of target populations: Definitions that are operationally useful for sampling and fieldwork purposes. For example, a definition that clearly identifies what constitutes a sex worker, in terms of duration of selling sex, form of payment, type of venue where they work, etc.

Operations manual: A document that describes every step to be taken during the implementation of a survey or study. Ideally, it provides standard operational procedures for every foreseeable occurrence.

Opportunistic infections: Illnesses caused by various organisms infecting immunodepressed persons that usually do not cause disease in persons with healthy immune systems. Persons with advanced HIV infection (that is, AIDS) suffer opportunistic illnesses of the lungs, brain, eyes, and other organs. These illnesses are referred to as AIDS-defining illnesses or conditions.

Opt-in: A patient or participant agrees to be tested.

Opt-out: A patient or participant refuses to be tested.

Optical density: The intensity of colour as measured by a machine in an EIA HIV antibody test, indicating whether the patient's sample is HIV-positive.

Ordinal variable: Variables that have a natural order, such as level of education.

Over-sampling: A sample may obtain more members of a particular sub-group than their representation in the target population warrants. In some cases, over-sampling is carried on purpose to learn more about a small sub-group, such as female injection drug users in communities that are predominantly male.

p24 antigen: A protein that appears in the serum of infected individuals approximately one week before HIV antibodies appear, or about 14 days after actual infection. In very large sero-surveys, persons who tested negative for HIV antibody can be retested for p24 antigen.

Pandemic: An epidemic occurring over a very wide geographic area (several countries or continents) and usually affecting a large proportion of the population. HIV is an example of a pandemic.

Parameter: The summary numerical description of variables about the target population.

Parenteral transmission: Transmission of an infectious agent through blood. Parenteral transmission of HIV can occur from the sharing of injection drug equipment, from transfusions with infected blood or blood products, or from needle stick injuries.

Participant observation: A qualitative research method in which direct observation is carried out over a period of time, and which is understood and accepted by the group being observed.

Participation bias: Error in results from a study that is due to differences in characteristics between those who participate in a survey and those who do not. For example, persons who already know they are HIV-infected may find testing unnecessary; those who suspect they are HIV-infected may decline testing in order to avoid stigma.

Partner concurrency: Having extensive sexual network connections to many persons at the same time, which increases the spread of HIV and STIs.

Passive surveillance: A system in which a health-care provider or worker notifies the health authority of any cases of these diseases, as opposed to ‘active surveillance.’

Pathogen: A biological agent that causes disease or illness to its host (for example, bacteria or virus).

Payment coupon: Kept by the recruiter. He/she will use it to claim an incentive for having recruited a peer into the study.

Perinatal transmission: Transmission of an infectious agent, such as HIV, from mother to baby before, during, or after the birth process. Also known as 'vertical transmission' or 'mother-to-child transmission.'

Period prevalence: The amount a particular disease that is present in a population over a specified period of time.

Pie chart: A circular chart in which the size of each 'slice' is proportional to the frequency of each category of a variable. A pie chart compares subclasses or categories to the whole class or category using different coloured slices.

PLACE: See 'Priorities for local AIDS control efforts.'

PLWHA: Acronym for 'Persons living with HIV/AIDS.'

PMTCT: Acronym for 'prevention of mother-to-child transmission.'

Point estimate: The amount of a particular disease present in a population.

Point prevalence: Refers to prevalence at a single point in time. Also known as 'point incidence.'

Population: The total number of inhabitants of a given area or country. In sampling, the population may refer to the unit from which the sample is drawn, not necessarily the total population of people.

Population-based sero-survey: A type of sero-survey that uses a probability sample of a population defined by geographic boundaries, such as villages or provinces, in order to obtain a direct measure of HIV prevalence in a general population.

Population sub-group: A group within a population that share certain characteristics or behaviours.

Positive controls: Specimens known to be positive, as used in proficiency testing.

Positive predictive value: The probability that a person with a positive test result is infected; in surveillance this refers to the proportion of cases reported by a surveillance system or classified by a case definition which are true cases. Also known as 'predictive value positive.'

PPS: See 'Probability proportional to size sampling.'

Precision: Refers to how well the results can be reproduced each time the survey is conducted.

Presumptive clinical diagnosis: Diagnosis made solely on the basis of symptoms, without the use of specific diagnostic tests.

Pre-surveillance assessment: Describes a set of activities that occur prior to beginning formal HIV and behavioural surveillance in *high-risk* groups. These activities include developing detailed plans and reviewing and collecting information that will help in planning and designing surveillance activities.

Prevalence: The proportion of persons in a given population with a disease or condition at a given point in time; a specific group infected. Prevalence is a direct measurement of the burden of disease in a population.

Prevalence assessment: Surveys that determine prevalence of a disease in a population.

Prevalence monitoring: Monitoring prevalence repeatedly over time to track trends.

Primary incentive: The incentive a participant gets for enrolling in the study and completing an interview.

Primary units: A sampling frame of larger unit. When it is difficult or impossible to make a list/sampling frame of each individual in the target population, we can develop a sampling frame of some larger unit; that is, clusters or primary sampling units. We then sample in stages by first sampling clusters and then sampling people within the clusters.

Priorities for Local AIDS Control Efforts (PLACE): A new, rapid assessment tool used to identify high transmission areas, which formalises the collection of information on high transmission areas. PLACE uses key informants to identify sites where people meet new sex partners, then interviews people at the site in order to characterise the site in each area and map sites, and, finally, interviews individuals socialising at the site to describe the characteristics of the people at the site.

Priority communicable disease: These are diseases that have the potential for epidemic spread and can be controlled through public health action. They are the diseases included in the Integrated Disease Surveillance form.

Prisoner: Any person involuntarily confined or detained in a penal institution, including persons detained pending arraignment, trial, or sentencing.

Probability proportional to size sampling: A class of unequal probability sampling in which the probability of a unit being sampled is proportional to the level of some known variable (Levy & Lemeshow).

Probability sampling: A sampling scheme that ensures that each entity in a population has a known, non-zero chance of being selected.

Process evaluation: An evaluation of a programme that determines how well the programme is functioning, as opposed to 'impact evaluation.'

Proficiency panel: A set of samples designed to judge the accuracy and precision of a laboratory. A necessary component of laboratory quality assurance. In the context of HIV testing this may be a group that contains approximately six HIV-negative and HIV-positive (weak to strong) specimens representative of the HIV strains circulating in a country and of the different stages of HIV infection. The panel should be sent to participating laboratories once or twice each year for quality assurance testing.

Proficiency testing: The act of sending a proficiency panel to a laboratory, designed to test the accuracy and precision of that laboratory.

Prophylaxis: Treatment to prevent or suppress infection, often given before a person's exposure to the pathogen. For example, the treatment given to mother's during childbirth in order to prevent infection of the newborn child.

Proportion: The relationship of a part to the whole, in which the numerator is included in the denominator; often depicted as a percent by multiplying by 100.

Prospective case reporting: To watch a group of cases for outcomes, such as the development and progress of HIV disease, over time and to relate this to other factors such as suspected risk or protection factors.

Prostitués homosexuels: Homosexual prostitutes. Male sex workers who identify as homosexual or gay.

Protocol: The detailed plan for conducting a research study or other activities in which specific steps are required, including surveillance activities.

Purposive sampling: A non-random sampling method that involves choosing respondents with certain characteristics.

Qualitative research: Research that focuses on the characteristics or quality of things, rather than the quantity. The sample included qualitative research is usually much less used than that included in quantitative research.

Quality assurance: The dynamic and ongoing process of monitoring a system for reproducibility and reliability of results that permits corrective action when established criteria are not met.

Quality control: A laboratory's internal processes for running specimens to ensure that the test equipment and reagents function properly.

Quantitative research: Research that focuses on quantity of things, rather than the quality. Quantitative research has powerful tools for the analysis of numbers, but researchers know that the things counted are often qualitative categories or definitions.

Questionnaire faults: Problems with the way questions are phrased, set out and ordered, which lead to misunderstandings of the questions.

Random error: Also called non-systematic error. This is the type of error that results from chance and leads to imprecise results.

Random sample: A sample derived by selecting individuals such that each individual has the same probability of selection.

Random walk: A variation of link-tracing sampling procedure in which the respondent is asked to give the names of other members of a hidden population. From that list, one is selected randomly, located and added to the sample. The process is repeated for a desired number of waves. (S.K. Thompson et al.)

Range: The difference between the largest and smallest values in a distribution.

Rapid assessment and response (RAR): A method that is used to assess the nature and extent of a public health problem and to suggest ways to address the problem. RAR is not designed as a surveillance tool, but as a way to assess a situation quickly, and bring in resources to address it.

Rapid HIV test: An HIV antibody test that is simple, does not require any reagents or equipment other than what is contained in the kit and provides results in less than 20 minutes.

Rapid plasma reagin test (RPR): A common serologic test for syphilis. Specifically, a non-treponemal test for anticardiolipin antibodies.

Rate: An expression of the frequency with which an event occurs in a defined population.

Ratio: The quantitative relationship between two or more things; the value obtained by dividing one quantity by another.

RDS: See ‘Respondent driven sampling.’

RDSAT: Acronym for respondent driven sampling analysis tool (a freeware software package for analysing RDS data).

Reference laboratory: A laboratory that functions as a recognised centre of expertise and standardisation of diagnostic techniques.

Referral coupon: Used by the recruiter to recruit a peer into the study.

Refugees: By legal definition, refugees are persons who are outside their country of nationality and who are unable or unwilling to return to that country. They cannot return due to a well-founded fear of persecution because of race, religion, political opinion or membership in an ethnic or social group.

Relative risk: A comparison of the risk of some health-related event such as disease or death in two groups. For example, an HIV-uninfected individual who has sexual intercourse with an HIV-infected person once a year may have a 5% chance of infection. But if the uninfected individual uses a condom every time, the relative risk when compared to condom non-use is 15%.

Reliability: Refers to how reproducible a result is from repeated applications of a measure to the same subject.

Representative sample: A sample whose characteristics correspond to those of the original population or reference population.

Representativeness: The degree to which the sample truly reflects the study population (that is, whether it is representative of the study population).

Resistance: The ability of an organism, such as HIV, to overcome the inhibitory effect of a drug.

Resource assessment: A component of RAR, a systematic examination of the response (funds, people, buildings, knowledge) that is either available or required to solve the problem.

Respondent driven sampling (RDS): A sampling technique that does not require a sampling frame. It is an adaptation of a non-probability sampling method (snowball sampling) and is based on the assumption that members of the sub-population themselves can most efficiently identify and encourage the participation in surveillance of other sub-group members. RDS starts with initial contacts or 'seeds' who are surveyed and then become recruiters. Each of these recruiters is given coupons to use to invite up to three eligible people that he/she knows in the high-risk group to be interviewed. The new recruits bring their coupon to a central place where they are interviewed. The recruits then become recruiters. This occurs for five to six waves. Both the recruits and the recruiters are given incentives to encourage participation.

Retrospective case reporting: To look backwards and examine exposures to disease, for example, HIV infection, and suspected risk or protection factors in relation to an outcome (infection) that is established at the start of the reporting.

Retrovirus: A type of RNA virus that produces reverse transcriptase which converts RNA into DNA. HIV is an example of a retrovirus.

Reverse-transcription: The process by which HIV's genetic material (RNA) is transformed into DNA, which allows it to fuse with the host's genetic material (DNA).

RIBA: Acronym for recombinant immunoblot assay, also known as Western blot. Immunoblot assays confirm anti-HCV reactivity. Serum is incubated on nitrocellulose strips on which four recombinant viral proteins are blotted. Color changes indicate that antibodies are adhering to the proteins. A positive result is if two or more proteins react and form bands. An indeterminate result is if only one positive band is detected.

Risk: The probability that an event will occur; for example, that an individual will become ill within a stated period of time.

Risk factor: An aspect of personal behaviour or lifestyle; an environmental exposure; an inborn, inherited, or demographic characteristic. Associated with an increased occurrence of disease or other health-related event or condition. For example, injection drug use is a risk factor for acquiring HIV.

RPR: See 'Rapid Plasma Reagin test.'

Safety protocol: A study document that describes how to deal with field incidents or adverse events.

Sample: A selected subset of a population. There are specific types of samples used in surveillance and epidemiology such as convenience, systematic, population-based and random.

Sample size: The number of subjects to be used in a given study.

Sample frame: A list of units from which a sample may be selected. A sample frame is a fundamental part of probability sampling.

Sampling bias: Also called selection bias. This refers to errors in sampling that decrease accuracy and lead to incorrect estimates. We also use the term 'biased samples' to mean that errors were made in choosing the people in the sample.

Sampling element: Individual member of the population whose characteristics are to be measured. See 'Sampling unit.'

Sampling error: The part of the total estimation error of a parameter caused by the random nature of sampling.

Sampling interval: The standard distance between elements selected in the sample population.

Sampling scheme: Procedure for choosing individuals to be included in a sample.

Sampling units: Refers to individual members of the population whose characteristics are to be measured. See 'Sampling element.'

Sampling variation: Difference between the estimate you measure in a sample and the true value of the variable in the study population.

Scale line graph: A graph that represents frequency distributions over time where the Y-axis represents frequency and the X-axis represents time

Second-generation surveillance*:* Built upon a country's existing data collection system, second-generation HIV surveillance systems are designed to be adapted and modified to meet the specific needs of differing epidemics. This form of surveillance aims to improve the quality and diversity of information sources by developing and implementing standard and rigorous study protocols, using appropriate methods and tools. Second generation surveillance refers to activities outside of those activities generally considered to be a part of routine case surveillance such as case reporting and sentinel sero-surveys and uses additional sources of data to gain additional understanding of the epidemic. It includes biological surveillance of HIV and other STIs, as well as systematic surveillance of the behaviours that spreads them.

Secondary incentive: The incentive a participant gets for recruiting his or her peers into the study.

Seeds: Non-randomly selected (by the investigators) members of the target population who will initiate the RDS recruitment process by recruiting members of his or her peer group. From each seed, a recruitment chain is expected to grow.

Selection bias: A systematic error in the process respondent selection for a study or survey.

Sensitivity: The proportion of persons with disease who are correctly identified by a screening test or case definition as having disease.

Sentinel case reporting: Reporting cases of a disease from sentinel sites.

Sentinel populations: Populations that are subject to sentinel surveillance activities. They may not necessarily be representative of the general population, but rather they might be the first affected by HIV. Examples include sexually transmitted infection patients or truck drivers.

Sentinel sites: Sites at which sentinel surveillance activities take place, including clinics attended by individuals who may or may not be representative of the general population but are likely to represent groups initially infected or at higher risk for infection than the general population.

Sentinel surveillance: A surveillance system in which a pre-arranged sample of reporting sources at 'watch post' or 'sentinel' sites agrees to report all cases of one or more notifiable conditions. Often designed to provide an early indication of changes in the level of disease. Depending on the nature of the population surveyed, these data may be representative of the general population, or they may simply give more detailed information about the populations tested.

Sero-conversion: The development of antibodies to a particular microbe. When people develop antibodies to HIV, they 'sero-convert' from HIV-negative to HIV-positive.

Sero-incidence surveillance: Collecting blood samples for measuring newly acquired HIV infection for the purposes of surveillance.

Serologic test: A blood test that determines the presence of antibodies to particles such as viruses. For example, a blood test that detects the presence of antibodies to HIV.

Sero-prevalence: The proportion of a population that is infected, as determined by testing blood for the appropriate antibody. For example, the proportion of a population that is infected with HIV, as determined by testing for HIV antibodies in blood samples.

Sero-prevalence surveillance: Collecting blood samples for the purpose of surveillance. Latent, sub-clinical infections and carrier states can thus be detected, in addition to clinically overt cases. This is especially important in the case of HIV and other STIs, which often have a long latent period before symptoms are apparent.

Sero-status: Refers to the presence/absence of antibodies in the blood. For example, the presence or absence of HIV.

Sero-surveillance: Collecting blood samples for the purpose of surveillance. Latent, sub-clinical infections and carrier states can thus be detected, in addition to clinically overt cases. This is especially important in the case of HIV and other STIs, which often have a long latent period before symptoms are apparent.

Sexual transmission: Transmission of an infectious agent, such as HIV, that occurs predominately through unprotected vaginal or anal intercourse, and less frequently through oral intercourse.

Sexually transmitted diseases: Symptomatic. Caused by organisms that are spread by sexual contact from person to person.

Sexually transmitted infection (STI): Asymptomatic. Diseases that are spread by the transfer of organisms from person to person during sexual contact.

Sex workers (SWs): Persons who engage in **sex work**, or the exchange of sex for money, which includes many practises and occurs in a variety of settings. These may include '**direct**' or '**formal**' sex workers, who are sometimes included in registries and

often found in brothels, and '**indirect**' or '**casual**' sex workers, who do not engage in sex work full time and are unlikely to be included in registries. The term 'sex worker' can be used to refer to female, male and transgendered sex workers.

Simple random sampling (SRS): Sampling where everyone has an equal chance of being randomly selected (a non-zero probability) and we know what that chance is.

Skewed: A distribution that is asymmetrical and does not follow a normal (bell-shaped) distribution.

Snowball sampling: Relies on informants to identify other relevant study participants in a chain referral pattern. Informants (seeds) who meet inclusion criteria are identified. This sampling design is based on chain referral and relies on the seed(s) to identify other relevant subjects for study inclusion. Those other subjects may identify other relevant subjects for inclusion. Snowball sampling is useful for studying populations that are difficult to identify or access. Representativeness is limited.

Social influence: Mild peer pressure from the recruiter who will receive a secondary incentive for recruiting his/her peers.

Social network: Members of a peer group who know each other.

Socio-metric stars: Seeds who are not only willing to recruit their peers, but are well-regarded by their peers and have a lot of them. Such seeds are more likely to influence others to be recruited into the study.

Specificity: The proportion of persons without disease who are correctly identified by a screening test or case definition as not having disease.

SRS: See simple random sampling.

Stacked bar chart: See 'clustered bar chart.'

Stakeholders (or stakeholder's group): Those with an interest in the results of surveillance activities. Includes public health practitioners, healthcare providers, data providers and users, representatives of affected communities; governments at the district, province and national levels; members of professional and private non-profit and donor organisations.

Standard error: Estimate of precision in probability sampling that can be used to construct a range of values within which the true population measure is likely to fall. We usually want to be 95% sure that the true population measure lies in our range.

Standardised Testing Algorithm for Recent HIV Sero-conversion (STARHS): A calculation for measuring new infection that uses a single blood test. STARHS uses the

results of two EIA tests, one highly sensitive and another modified to be less sensitive. The less sensitive EIA test is called the 'detuned' assay.

Statistics: A branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate population parameters.

Steering method: In RDS, using additional methods to recruit a special sub-population of interest; for example, providing an extra coupon to be used only to recruit female IDUs.

STI: See 'sexually transmitted infection.'

Stigma: A mark of disgrace or shame. For example, in some societies, being infected with HIV causes a person to be stigmatised.

Strata: A sub-group in stratified sampling.

Strategic information (SI): Refers to any data collected by surveillance or monitoring and evaluation of a programme or system. Includes, but is not limited to, process indicators, output indicators and surveillance data.

Stratification: The classification of a survey population into sub-groups or strata on the basis of selected characteristics.

Stratified and constant incentives: In a study of SWs, a constant incentive level was considered too low to attract the more hidden SWs who earned a higher income. The research team considered using a stratified incentive process. The SWs received an incentive based on the type of sex work they did. For instance, a street-based SW received a $5.00 incentive, while a call-girl-type SW received a $10.00 incentive

Stratified sampling: Stratified sampling is generally used to obtain a representative sample when the population is heterogeneous, or dissimilar, where certain homogeneous, or similar, sub-populations can be isolated (strata). A stratified sample is obtained by taking samples from each stratum or sub-group of a population.

Street children: Children who live and/or work on the streets, including orphaned, homeless, runaway, or neglected children who live chiefly in the streets without adequate protection, supervision, or direction from responsible adults.

Subcutaneously: Below the skin, as in an injection.

Sub-population: See 'population sub-group.'

Sufficient cause: A causal factor or collection of factors whose presence is always followed by the occurrence of the effect (of disease).

Surveillance: The systematic collection, analysis, interpretation, and dissemination of health data on an ongoing basis, to gain knowledge of the pattern of disease occurrence and potential in a community, in order to control and prevent disease in the community.

Surveillance sites: The places from which case reports are obtained. This includes sites at which universal reporting and sentinel reporting are done. These may be healthcare facilities or other locations at which sero-surveys are conducted.

Survey population: The target population modified to take into account practical considerations (for example, all commercial sex workers in a city over the age of 15, excluding those who are based at home, as they cannot be accessed).

Survey protocol: A manual that describes all the steps and tasks involved in a sero-survey.

Survival sex: To barter sex for the necessities of living, such as food, shelter, goods, money. Engaged in by vulnerable populations, for example, by displaced women, street children, and transgendered people who are marginalised and discriminated against.

Susceptible: Vulnerable or predisposed to a disease.

Symptomatic: Exhibiting symptoms.

Symptoms: Any perceptible, subjective change in the body or its functions that indicates disease or phases of disease, as reported by the patient.

Syndrome: A group of symptoms as reported by the patient and signs as detected in an examination that together are characteristic of a specific condition.

Syndromic case reporting: A surveillance system in which a diagnosis of the infection is made through the presence of symptoms using a standard case definition. Frequently used for surveillance of sexually transmitted infections in countries in which access to laboratory testing may be limited.

Syndromic prevalence: The prevalence of a particular syndrome, or set of symptoms, in a given population. Usually calculated when testing equipment is not available to verify the presence of particular pathogen in a laboratory.

Syphilis: A sexually transmitted disease resulting from infection with the bacterium *Treponema pallidum*. Syphilis can also be acquired by newborns from their mothers during pregnancy.

Systematic sampling: A sampling method that consists of randomly selecting the initial patient who meets the inclusion criteria and then selecting every ‘nth’ (for example, third or fifth) eligible patient thereafter until the predetermined sample size is reached or the survey period is over.

Systemic: Concerning or affecting the body as a whole.

Table: A set of data arranged in rows and columns.

Target population: The group that meets a survey's measurement objective (for example, all commercial sex workers in a city).

Targeted sampling: Targeted sampling uses pre-existing indicator data (qualitative and quantitative) to construct a sampling frame from which recruitment sites are then randomly selected. Qualitative indicator data includes ethnographic data and key informant interviews. Types of quantitative indicator data include cases of HIV/AIDS and STIs, admissions to drug treatment and population characteristics from census data. There are several limitations: 1) indicator data may not be useful in characterising the target population; 2) sampling may be biased and difficult to replicate; 3) geographic areas may not be sampled in proportion to the number of members in the population of interest; 4) the population of interest may not be sampled in proportion to the intensity of risk behaviour and 5) the probability of selecting a member of the population of interest may not be known.

TB: Acronym for tuberculosis.

Testing (HIV) strategy: The use of an appropriate HIV test or combination of HIV tests. The choice of testing strategy used is based on the objective of the test, the sensitivity and specificity of the test, and HIV prevalence in the population being tested.

T-helper lymphocyte: Also known as 'T-cell.' Immune cells that seek and attack invading organisms. HIV enters T-cells through their CD4 receptor proteins, making T-cells virtual HIV-factories.

Time-location sampling (TLS): Similar to conventional cluster sampling, but gets around the problem of clusters that are not stable (that is, clusters where the number and type of people vary by, for example, time of day). Time-location sampling allows the same site to be included in the sample frame more than once (for example, at different times of the day or different days of the week).

Timeliness of reporting: One of several attributes of a surveillance system. Timeliness may be defined as the time period between the diagnosis of the disease and the receipt of a case report form at the health district.

Transactional sex: Distinct from other forms of commercial sex. Includes the receipt of gifts or services in exchange for sex.

Transgendered persons: Persons who identify with or express a gender and/or sex different from their biologic sex.

Transition probability: The likelihood that a person will change from one state to another, for example becoming HIV positive.

Transmission: Any mode or mechanism by which an infectious agent is spread through the environment or to another person.

Trend: A long-term movement or change in frequency, usually upwards or downwards.

Treponema pallidum: The bacterial causative agent of syphilis.

Triangulation: The process of examining several different sets of data, which are measuring different things to come up with a better understanding of how and where an epidemic is spreading. For example, the use of antenatal clinic data, census data, and registered deaths in order to create a more complete picture of the AIDS burden in a country.

***Trichomonas vaginalis*:** A sexually transmitted protozoan parasite that causes the vaginal infection, **trichomoniasis**, characterised by itching, burning and vaginal discharge. Reinfection is common if sexual partners are not treated simultaneously.

True negatives: Test results that are negative when the patient actually does not have the disease that is being tested for.

True positives: Test results that are positive when the patient actually has the disease that is being tested for.

Tuberculosis: An airborne, often fatal bacterial infection caused by *Mycobacterium tuberculosis.* It causes damage to the lungs and other parts of the body. Infection is more likely in people with weak immune systems.

UAT: See 'unlinked anonymous testing.'

UNAIDS: Acronym for The Joint United Nations Programme on HIV/AIDS.

UNGASS: Acronym for United Nations General Assembly Special Session on HIV/AIDS.

Univariate analysis: The most basic, yet often the most important, type of behavioural surveillance analysis, because it shows the distribution of each variable. Most of the indicators defined for behavioural surveillance purposes are calculated through univariate analysis. They would include variables like the proportion of young men who have had sex with more than one partner during a given time period. When trends are analysed, statistical techniques are used to calculate how likely it is that changes in the proportions could have occurred by chance, or whether observed changes are likely to reflect real changes.

Universal case reporting: A surveillance system in which all persons who are identified as meeting the case definition for a particular disease are reported. For example, all persons with AIDS who receive care at any healthcare facility are reported. This is in contrast to sentinel reporting in which only selected sentinel sites report all persons who meet the case definition.

Universal conscription: Military conscription in which all physically able men between certain ages (for example 17-28) must perform military service.

Universal precautions: Recommendations issued by CDC to minimise the risk of transmission of bloodborne pathogens, particularly HIV and HBV, by healthcare and public safety workers. Barrier precautions are to be used to prevent exposure to blood and certain body fluids of all patients.

Unlinked anonymous testing (UAT): Testing that occurs when a sample of blood originally collected for other purposes is tested for HIV after being anonymised. The person whose blood is taken does not know that his/her blood will be tested for HIV. All information that could identify the person is removed from the sample so that the results of the test cannot be linked back to them.

Unprotected sex: Having sex without using a condom as protection against HIV and other sexually transmitted infections.

Urethritis: Inflammation of the urethra.

Vaccine: When injected into an individual, a vaccine protects against subsequent infection by a particular organism or results in a less severe illness should infection occur. Currently there is no vaccine for HIV.

Validity: The validity of a measure is the extent to which it actually measures what it is suppose to measure: the truth.

Values: Magnitude of measurements (statistics).

Variable: Any characteristic or attribute that can be measured.

VCT: See ‘voluntary counselling and testing.’

VDRL: See ‘Venereal Disease Research Laboratory test.’

Venue-based: Locations in the community, such as bars, tea houses, and street corners.

Venue-based sampling: Recruit respondents in places and at times where they would reasonably be expected to gather. The venues act as screeners in identifying potential respondents. Venue-based sampling requires comprehensive formative research.

Venereal Disease Research Laboratory test (VDRL): A common serologic test for syphilis. Specifically, a non-treponemal test for anticardiolipin antibodies.

Vertical surveillance system: See 'categorical surveillance system.'

Vertical transmission: See 'perinatal transmission.'

Vesicular: Pertaining to vesicles or blisters.

Viral load: The amount of HIV in the circulating blood. Also known as 'viral burden' or 'viral dose.'

Viral load test: Test that measures the quantity of HIV in the blood.

Virulence: The relative capacity of an organism to overcome the body's immune defences.

Virus: Micro-organisms that typically contain a protein coat surrounding nucleic acid (RNA or DNA) that are capable of growth only within living cells.

Vital records: Certificates of birth, death, marriage and divorce that are required by law.

Voluntary counselling and testing (VCT): A programme that provides both counselling and testing services to communities, allowing persons who are tested to obtain emotional and medical support before and after their HIV tests.

Voluntary migrants: People who temporarily work or travel away from their homes.

Volunteerism: A term used to describe overly cooperative subjects, leading to a potential bias if such cooperative people differ from the rest of the population of interest.

Vulnerable population: A group whose members are discriminated against and who face stigma, making them vulnerable to negative consequences of surveillance, including social and physical harm.

Western blot: A type of HIV test, Western blot uses an electroblotting method in which proteins are transferred from a gel to a thin, rigid support and detected by binding of labeled antibody to HIV.

WHO: Acronym for the 'World Health Organization.'

Width: See 'interval width.'

X-axis: The horizontal line of a graph, usually found at the bottom.

Y-axis: The vertical line of a graph, usually found at the left but sometimes also at the right.

Years of potential life lost: A measure of the impact of premature mortality on a population, calculated as the sum of the differences between some predetermined minimum or desired life span and the age of death for individuals who died earlier than that predetermined age.

YLL: See ‘years of potential life lost.’

Notes

Appendix C, Useful Links

Family Health International (FHI)
Family Health International has pioneered ways to curtail the spread of HIV/AIDS. Many of the HIV prevention "best practices" in use today have emerged from FHI's work in more than 60 countries.
www.fhi.org/en/HIVAIDS

HIV/AIDS Survey Indicators Database
The HIV/AIDS Survey Indicators Database is overseen by a technical advisory committee that includes representatives from USAID, UNICEF, CDC, UNAIDS, WHO, US Census Bureau, Family Health International, MEASURE Evaluation, The Synergy Project, and MEASURE DHS+ (the implementing organisation). USAID is currently the primary funder for the initiative, with UNAIDS and UNICEF providing additional support. There are 180 surveys available in the database.
www.measuredhs.com

Multiple Indicator Cluster Survey (MICS), UNICEF
The Multiple Indicator Cluster Survey (MICS) is a household survey programme developed by UNICEF to assist countries in filling data gaps for monitoring the situation of children and women. It is capable of producing statistically sound, internationally comparable estimates of these indicators.
www.childinfo.org

Respondent-Driven Sampling (Cornell)
Defines RDS and provides information on minimum data requirements, sampling references, intervention references and downloads.
http://www.respondentdrivensampling.org

Respondent-Driven Sampling Field Experiences Message Board
A respondent-driven sampling site where people can share questions and receive answers in real time. The Board is not monitored so it relies on people checking in for now. It requires registration. You will be alerted to new postings if you choose. Feel free to share the web site with other RDS users or others who might be interested.
http://www.globalhivevaluation.org/messageboard/topicView.aspx?id=2

UNAIDS (Joint United Nations Programme on HIV/AIDS)
As the main advocate for global action on HIV/AIDS, UNAIDS leads, strengthens and supports an expanded response aimed at preventing the transmission of HIV, providing care and support, reducing the vulnerability of individuals and communities to HIV/AIDS and alleviating the impact of the epidemic. UNAIDS compiles epidemiological fact sheets about each country involved in HIV/AIDS prevention programs, as well as specific populations.
www.unaids.org

United Nations General Assembly Special Session on HIV/AIDS (UNGASS)
The United Nations General Assembly Special Session on HIV/AIDS (UNGASS) has developed a set of core indicators. Monitoring the Declaration of Commitment on HIV/AIDS Guidelines on Construction of Core Indicators is available online at: http://www.ungass.org/index.php/ungass/ungass/meeting_ungass_targets/ungass_core_indicators

United States Department of Commerce, U.S. Census Bureau's International Programs Center
The International Programs Center, part of the Population Division of the U.S. Bureau of the Census, conducts demographic and socio-economic studies and strengthens statistical development around the world through technical assistance, training, and software products. The IPS maintains an HIV/AIDS Surveillance database, the Monitoring the AIDS Pandemic (MAP) Network, and a series of HIV/AIDS country profiles. The Programs Center provides various country profiles that examine the patterns and trends of the epidemic, as well as maps and tables that serve to summarise the statistics for each region in a streamlined format.
www.census.gov/ipc/www/hivaidsn.html

Appendix D

Answers to Warm-Up Questions and Case Studies

Answers are provided in *italics* for each unit's warm-up questions and case study.

Answers to the questions within the unit are not included. Unit questions are designed to stimulate small group discussion among participants in the workshop or class.

Unit 1 Answers

Warm-up questions

1. True or false? One-time cross-sectional surveys can be considered surveillance. *False. Although they are often the beginning of a surveillance system, one-time cross-sectional surveys should not be considered surveillance.*

2. _____surveillance involves regular, repeated cross-sectional surveys collecting data that can be compared over time on HIV risk behaviours and other relevant issues.

   a. *behavioural*
   b. biological

3. Which of the following is a use of behavioural surveillance?

   a. to explain changes in HIV prevalence over time
   b. to provide information for prevention programmes
   c. to raise the awareness of HIV among policy-makers
   d. to provide an early warning for areas and groups at risk for transmission of infection
   e. *all of the above*

4. True or false? Surveillance is a useful tool for evaluating specific HIV/AIDS interventions. *False. Without adaptation, surveillance only provides evidence for the impact of HIV programmes as a whole, and not for the impact of specific interventions or specific programme elements.*

5. True or false? In a generalised epidemic everyone is at equal risk of infection. *False. Even in a generalised epidemic, not everyone in the population is at equal risk of HIV, or has an equal role in the spread or maintenance of the epidemic.*

6. _____ *sites* are facilities such as STD clinics, antenatal care clinics, blood donation centres, drug treatment programmes, prisons and needle exchange programs
   a. *sentinel*
   b. community

7. Which of the following is the definition of linking behavioural and biological data?
   a. collecting HIV, STI and behavioural data from the same individuals at the same time
   b. collecting HIV, STI and behavioural data from the same source population at different times
   c. analysing HIV, STI and behavioural data from similar source population, using whatever data are available
   d. reporting behavioural and biological surveillance together
   e. *all of the above*

8. Collecting _____ level data provides more detailed information but requires larger sample sizes and, therefore, more time and money
   a. national
   b. *sub-national*

Case study

1. When discussing guidelines for whom to include in surveillance for each epidemic state, define what is meant by the "general population" and by "high-risk group surveillance"?

The general population is the people living in a region or nation. High-risk groups are those whose behaviours, occupations or lifestyles could expose them to higher risk of acquiring and transmitting HIV than the rest of the population. These groups are often important in establishing, accelerating or sustaining the HIV epidemic. Remember that high-risk groups are part of the general population. They have links with other population groups and membership can overlap. For example, commercial sex workers (CSWs) can also be intravenous drug users (IDUs). Ultimately, what is considered a high-risk group must be based on the local context.

2. Name a group you consider high risk. Why do you consider them high risk? Describe the group and tell why you think it is high risk.

Local information must be used to determine what constitutes a risk group based on a group's importance or potential importance in the epidemic. Most people view CSW as a risk group, but one could argue that CSWs are not actually a group. This could be argued because CSW can be sub-divided into many groups, and because individual women on the street who solicit sex do not always belong to or associate with any groups. Alternatively, people with multiple partners could be considered a 'group.'

Unit 2 Answers

Warm-up questions

1. What are two characteristics of a good indicator? *A good indicator measures something of relevance to the topic (the measure serves some use), measures the item of interest accurately, is easy to interpret and is defined in clear terms. It can be compared across different population groups and across time and is feasible to collect in terms of effort/cost.*

2. True or false? Behavioural surveillance indicators should measure aspects of behaviours that are integral components in the spread of HIV. *True.*

3. When should indicators be selected during behavioural surveillance?
   a. *during planning*
   b. during analysis

4. True or false? The two most difficult issues when creating behavioural surveillance indicators are defining the behaviours themselves and defining the time period for which the indicator should refer. *True.*

5. True or false? Behavioural surveillance indicators do not need to be consistent over time. *False*

6. The time reference period of an indicator should be determined by its _____.
   a. prevalence
   b. *frequency*

Case study

1. What do the following indicators measure? Which is more useful?

   a. Married men who go to a sex worker - *The first indicator gives us an idea about the size of the link between CSWs and married men. If a large proportion of married men use CSWs, commercial sex could potentially be important for driving the epidemic.*
   b. Men who go to sex workers who are married - *The second indicator does not tell us anything about the role of CSW in driving the epidemic. We do not know if the use of CSWs is high only in that, among those people who use CSW, a certain proportion are married.*

2. Discuss the differences between substance users and illicit drug users. *"Substance" can include alcohol and tobacco.*

Unit 3 Answers

Warm-up questions

1. What is measurement error? W*hen the data collected do not accurately measure the characteristics of interest, which affects the validity of our data.*

2. In observational studies, which of the following is a source of measurement error?

   a. interviewer error
   b. respondent error
   c. questionnaire faults
   d. *all of the above*

3. True or false? Bias can be controlled for during data analysis. *False.*

4. True or false? A face-to-face interview is the best data collection method. *False. Each data collection method has advantages and disadvantages, and there is no conclusive evidence that any one method is better, over-all, than another.*

5. Which of the following is an advantage of using the self-administered data collection method?

   a. *inexpensive to administer*
   b. no literacy requirement
   c. date entry step eliminated

6. Which of the following are advantages of adapting survey questions from surveys that have already been successfully implemented?

   a. *builds on current best practice of how questions can be best expressed*
   b. *saves time and money*
   c. *ensures consistency with other available data sources*
   d. eliminates the need to pre-test the questionnaire
   e. all of the above

9. What is one solution to the problem of interviewer safety when working with hard-to-reach groups?

   Select interviewers who know the area; Interviewers should work in pairs.

8. True or false? When having difficulty reaching members from hard-to-reach groups to show up for an interview, one solution is to use incentives. *True.*

Case study

1. What topics need to be taught during interviewer training? Outline a curriculum.

   - *Introduction to the purpose of the study*
   - *Rapport building and communication skills*
   - *Cultural sensitivity/knowledge of the target population*
   - *How to talk about sex with ease (how to be non-judgmental)*
   - *Survey protocol*
   - *Handling and tracking non-response*
   - *Controlling the interview, handling interruptions and respondent fatigue*
   - *Use of survey instruments—skip patterns, coding, phrasing and meaning, note-taking*

2. What topics need to be taught during supervisor training? Outline a curriculum.

   - *All of the interviewer training*
   - *How to respond to interviewee questions*
     Quality assurance within interviews
     Ethics: respecting confidentiality, protecting respondents
     Supervising and supporting fieldworkers
   - *Sampling methodology*
   - *Importance of quality control and standardisation*
   - *Awareness of falsified responses*
   - *Logistics in fieldwork*
   - *Safety of fieldworkers and respondents*
   - *Management skills*
   - *Ability to report to upper management*

Unit 4 Answers

Warm-up questions

1. The _____ is the group that meets a survey's measurement objective (for example, all commercial sex workers in a city).

   a. *target population*
   b. survey population

2. Is drawing names randomly out of a hat for sampling an example of probability sampling or non-probability sampling?

   a. *probability sampling*
   b. non-probability sampling

3. True or false? Non-probability sampling is prone to selection bias. *True.*

4. We can increase the precision (that is, decrease standard error) of our estimate by increasing the _____.

   a. *sample size*
   b. quality of interviewer training

5. What is an estimate of precision that can be used to construct a range of values within which the true population measure is likely to fall?

   a. *standard error*
   b. systematic sample

6. Which of the following is not a type of sampling
   a. stratified
   b. cluster
   c. respondent driven
   d. systematic
   e. *salient*

7. True or false? Statistical packages assume simple random sampling when performing statistical tests. *True.*

8. True or false? Cluster sampling provides more precise estimates of indicators than simple random sampling. *False. Cluster sampling results in a less precise estimate of our indicators than simple random sampling, as respondents within clusters may be similar to each other. We need to compensate for this by increasing the sample size.*

9. Which of the following is *not* a way to overcome sampling challenges for behavioural surveillance?

    a. using different sampling strategies for different groups
    b. *using convenience sampling when possible*
    c. using conventional sampling methods in unconventional ways
    d. using sampling techniques such as respondent-driven sampling (RDS)

Case study

For each of the following groups, decide:

- What is the best sampling strategy
- Why this is the best strategy?
- What are the strong and weak points of using this method for the group?

Group 1: Youth. Conventional cluster sampling would be appropriate in most settings.

Advantages of cluster sampling:

- *only need sample frame of clusters and individuals in selected clusters*
- *sample is concentrated in geographical areas*

Disadvantages of cluster sampling

- *decreases precision of estimate, requiring a larger sample size*
- *PPS sampling or weighted sampling required for unbiased estimates*

Group 2: MSM. Either TLS or RDS would be appropriate in most settings.

Advantages of TLS:

- *allows us to do a probability sample of populations that are hidden or floating*
- *only need sample frame of clusters and individuals in selected clusters*
- *sample can be concentrated in geographical areas*

Disadvantages of TLS:

- *mapping and ethnographic work can be time consuming and clusters/sites can close rapidly*
- *only reaches sub-set of population that come into contact with the locations where the sampling is done*
- *difficult to identify and access respondents*
- *difficult to maintain randomness while selecting respondents within clusters*
- *PPS not often done due to difficulties estimating cluster size, and samples often require weighting, which if not done can result in biased estimates.*

Advantages of RDS:

- *ease of field operations*
- *no need for ethnographic mapping or sampling frame*
- *target population recruits for you, which is useful when the group does not trust the research community*
- *less visible members of the population reached*
- *lower cost*

Disadvantage of RDS:

- *population must be a network*
- *need to keep track of links between recruiters and recruits*
- *ethical issues involved in using incentives*

Using Table 4.5, find the sample size you would need to do a respondent-driven sampling (RDS) survey of MSM. A survey of MSM in your city found that 80% of MSM could correctly name the ways that HIV is transmitted and ways that HIV can be prevented. Health education campaigns are planned for the coming year using MSM peer educators. You wish to show that HIV/AIDS knowledge improved by at least 10% by the following year. How many MSM do you need to include in this year's survey and next year's survey? Assume that knowledge may increase or decrease and you wish to have 80% power. What sample size would you need if you decided to use time-location sampling (TLS) instead of RDS?

The same sample size would be needed. From table 4.5, go down to .80 in the first column and .90 in the second column. The third column of 395 is the sample size needed for each year's survey if you use RDS. If TLS is used, then the sample size is also 395 per year. The sample sizes are the same because the design effect is the same.

Unit 5 Answers

Warm-up questions

1. True or false? It is better to share over-all responsibility between different staff for the data management. *False. It is usually better for one person to have over-all responsibility for the data than to share responsibility between different staff.*

2. Data management does not include which of the following?
   a. data coding
   b. data entry
   c. data cleaning and checking
   d. *data framing*

3. True or false? The data manager should not be involved during the questionnaire design. *False. The data manager should be involved during the questionnaire design stage, not only to review the questions, but to review the way in which the questionnaire is organised.*

4. Which type of behavioural surveillance analysis is performed to determine whether one variable is related to the distribution of another? (For example, an association between a respondent's age and his or her use of condoms.)

   a. univarite
   b. *bivariate*
   c. multivariate

5. Most of the indicators defined for behavioural surveillance purposes are calculated through _____ analysis.

   a. *univarite*
   b. bivariate
   c. multivariate

6. When data are presented, they should be packaged appropriately for the different _____.

   a. *audiences*
   b. data collection sequence

7. True or false? The surveillance cycle ends when the official report is published. *False.*

8. True or false? Data from other sources, including other countries, can be used to fill in any important gaps, as long as the source is made very clear. *True.*

Case study

How do you think data would be best presented to a senior politician in terms of language, length and messenger?

In communicating with policy-makers, pare your message to the bare essentials and tell them exactly what action they need to take. Formal pathways may need to be followed when communicating with policy-makers. You should also use all the informal networks you have, as getting through the front door can be difficult. Politicians and policy-makers, more than any other group, require the right messenger. Very senior policy-makers tend to have rather a small circle of senior advisers. These people may be easier to approach than the key policy-makers themselves, and they usually make very good messengers. Other channels that can be used include senior officials from international organisations or senior diplomats from donor countries.

Unit 6 Answers

Warm-up questions

1. Match each ethical principle with its definition.

__*b*__	respect for persons	a. Refers to minimising risk to individuals, not only physical risk but also risk of psychological harm and stigmatisation
__*a*__	beneficence	b. Requires investigators to see study subjects not as passive sources of data, but as persons whose rights and welfare must be protected
__*c*__	justice	c. Risks and benefits from studies should be distributed fairly and evenly in populations

2. It is important that everyone the surveillance system encounters is treated with *respect*, from community leaders and local officials, to those surveyed.

3. What is informed consent? *Informed consent means that you tell the person enough about the nature of the surveillance for them to make a proper (informed) decision about whether or not to take part.*

4. Name two pieces of information that should be provided before a person can make an informed decision to take part in a survey?

   - *the nature of the survey (for example, who is conducting the survey, purpose of the survey, length of interview, type of question, etc.)*
   - *the potential risks and benefits*
   - *how the information will be used*
   - *how their privacy will be protected (names or addresses are not written)*
   - *that participation is voluntary*
   - *participants have the right to refuse to answer any questions or stop the interview at any time, especially as they may find some of the questions sensitive*

5. True or false? In surveillance, written consent forms are the most appropriate way to document that the process of informed consent has occurred. *False. In order to ensure total confidentiality, it is usually best to obtain verbal consent, as this means that the name of the respondent does not need to be recorded.*

6. What are two ways to ensure a participant's confidentiality?

   - *ensure names or other means of identification are not recorded on surveys*
   - *store data safely and appropriately*
   - *train fieldworkers on the importance of confidentiality*
   - *have clear disciplinary procedures for staff who breach confidentially*
   - *identify problems and possible solutions related to confidentiality.*

7. Match each of the following ethical issues with a potential solution.

__b__	loss of earnings	a. ensure fully informed consent.
__a__	increases discrimination of the group	b. conduct interviews outside work times
__c__	participants get no direct benefit from surveillance	c. do not foster false expectations

Case study

Design a consent form to be used with female sex workers.

Below is an example of a consent form that could be used with FSW:

Introduction: "My name is_____. I'm working for_____. We're interviewing people here in [name of city, region or site] in order to find out about [describe purpose of study]. Have you been interviewed in the past few weeks [or other appropriate time period] for this study? **If the respondent has been interviewed previously during this round of BSS, do not interview the person again.** Tell them you cannot interview them a second time, thank them, and end the interview. If they have not been interviewed before, continue:

Confidentiality and consent: "I'm going to ask you some very personal questions that some people find difficult to answer. Your answers are completely confidential. Your name will not be written on this form, and will never be used in connection with any of the information you tell me. You do not have to answer any questions that you do not want to answer, and you may end this interview at any time. However, your honest answers to these questions will help us better understand what people think, say and do about certain kinds of behaviours. We would greatly appreciate your help in responding to this survey. The survey will take about XX minutes to ask the questions. Would you be willing to participate?"

_____

(Signature of interviewer certifying that informed consent has been given verbally by respondent)

Unit 7 Answers

Warm-up questions

1. Which of the following is an example of a question you should answer in order to select appropriate geographical areas during the pre-surveillance assessment?

   a. What are the regional differences in terms of HIV transmission and risk behaviour?
   b. Where are the interventions located?
   c. How much money and staff are available for surveillance?
   d. *All of the above.*

2. Match each pre-surveillance assessment method with its description.

__*a*__	Assess what is currently known about the national epidemic or sub-epidemic.	a. This step involves reviewing existing surveillance data, published and unpublished literature, and talking to people who are knowledgeable about the epidemic.
__*c*__	Conduct pre-surveillance assessment using qualitative assessment methods and mapping.	b. You can perform this step using existing information from general population surveys or collecting the information if none exists.
__*b*__	Gather information on risk behaviours and HIV levels in the general populations.	c. This step involves fieldwork to further identify and verify hotspots, as well gathering information to clearly define populations to be included in surveillance and to guide surveillance fieldwork.

3. Methods used in *quantitative* research include large sample size, random samples and shorter interviews. Methods used in *qualitative* research include unstructured questionnaires, lengthy interviews and fewer well-trained fieldworkers.

4. Suppose you want to determine in which areas of a city street-based sex occurs. Which of the following is not an appropriate method to use during a pre-surveillance assessment?

    a. literature/data review
    b. in-depth qualitative interviews
    c. general population survey
    d. *all are appropriate*

Case study

1. Make a list of all the things that need to be done before data collection begins.

Below are examples of some of the activities that must be done prior to data collection:

- initial consensus building meeting with surveillance committee
- identify data needs and key indicators
- identify and select populations to be included in surveillance and define eligibility criteria
- select the regions/cities/areas that will be included in the surveillance
- determine the sampling options, sampling frame and sample size requirements
- define eligibility criteria
- create alliances with the survey populations and gatekeepers
- prepare logistics and supplies
- design, revise, translate and pre-test the questionnaires
- identify potential fieldwork problems and solutions
- produce field plan including how quality will be ensured
- define the timeline, budget and available resources
- get ethical clearance
- recruit statistician and data managers
- produce a data analysis plan and develop data base
- produce a data use plan
- recruit and train supervisors, interviewers and data entry clerks

2. What do you want to know before writing a protocol? How will you find out? Discuss fully.

You will want to have background information on the epidemic in the region you are sampling. You will also need to decide what population you will be sampling and what geographic area you will be sampling from. Lastly, you will need to determine what sampling strategy will be most effective to use in your population and region of interest.

In order to find out this information, you should assess what is currently known about the national epidemic or sub-epidemic, conduct a pre-surveillance assessment using qualitative methods and mapping, and gather information on risk behaviours and HIV levels in the population.

Final Case Study

**The following questions are intended to be open-ended and can have several different responses. Below is just one example of potential responses to the questions.*

Part I

Global campaigns have focused on the fact that 50% of new HIV infections occur in people under the age of 24—an age at which people are discovering their sexuality and likely to be engaging in risk behaviour.

A country with a rapidly growing epidemic pooled their national surveillance data and analysed them by age. It was found that some 68% of all existing infections were among young people under 25.

A front-page article in the leading newspaper reports that behavioural surveillance among secondary school children had turned up shocking indicators: over half of the students had multiple sex partners and only 20% used condoms.[4]

As a ministry of health official:

- What specific programme(s) might you implement in response to these data?
  - *school-based programmes*
  - *health education*
  - *peer educators*
  - *condom promotion*
  - *abstinence/partner reduction*
- To what population(s) will you target your programs?
  You will want to target the program towards young people under 25
- What additional information do you need to make better programming decisions?
  - *the primary mode of transmission for recent HIV infections*
  - *geographic distribution of HIV infections*
  - *among young adults, any subpopulations with elevated risk of HIV*
  - *proportion of sexually active young adults*
  - *proportion of young adults with multiple sex partners*
  - *other behavioural risk factors for transmission of HIV*

[4] Extract from the data-use module: Using data from HIV surveillance systems: Guidance on effective use of data from second generation surveillance systems: Elizabeth Pisani.

Part II

The Ministry of Health reacts rapidly, successfully lobbying for an allocation of US $4 million to develop life-skills curriculum focusing on helping young people to avoid unsafe sex.

The programme is implemented quickly, and two years later behavioural surveillance is repeated. The good news is that multiple partnerships among youth dropped from 50% to 25% and reported condom use doubles. However, the number of infections in young people reported by the HIV surveillance system has continued to grow.[5]

As a Ministry of Health official:

- How will you explain these data? What might account for them?
    - *Data quality could be an issue; check for accuracy*
    - *Multiple partnerships may not be driving HIV transmission*
    - *Other factors may be contributing to spread of HIV*
- What does this information tell you about your current HIV/AIDS programming? Would you make any changes?

  If the data is accurate, this information would tell us that the current HIV/AIDS programming is not effective in reducing HIV transmission in young people under 25. There may be other behavioural characteristics that are driving the epidemic. You should consider conducting further analyses of the surveillance data in order to look more closely at other behavioural risk factors.

[5] Extract from the data-use module: Using data from HIV surveillance systems: Guidance on effective use of data from second generation surveillance systems: Elizabeth Pisani

Part III

A parliamentary committee orders an inquiry. Public health officials go back to the behavioural surveillance data. It turns out that the indicators reported in the baseline year were accurate, in that half of all students who had had sex in the previous year had multiple partners and only 20% had used condoms. However, closer inspection of the data revealed that:

- Only 8% of all students had ever had sex at all, and only 4% had been sexually active in the past year
- 10% of all secondary school students had reported injecting drugs in the previous year and by the second round of behavioural surveillance that proportion had increased to 14%

These very high rates of injecting drug use had been overlooked because health officials, the press, and other agencies were looking for evidence of unprotected sex.

The ministry of health quickly ordered an assessment of injecting drug use and found that 85% of injectors were under the age of 25. Life-skills programs were subsequently redesigned to focus more on helping young people stay away from drugs. Harm reduction programmes were redesigned for those already injecting drugs, and injecting drug users were added as a group to the national surveillance system.[6]

What specific lessons can be drawn from this example?

Several lessons can be drawn from this example:

- *Programming decisions should be evidence-based.*
- *It is important not to jump to conclusions about reported data before carefully considering various explanations.*
- *Thorough data analysis, including subgroup analysis, is necessary to understand trends in the data.*
- *Complementing quantitative findings with qualitative research may help to better understand observed trends in the data.*

[6] Extract from the data-use module: Using data from HIV surveillance systems: Guidance on effective use of data from second generation surveillance systems: Elizabeth Pisani