

Methodological Annex

This annex provides some basic guidance on the issues of collecting data for monitoring and evaluation. Many of the issues discussed here are pertinent to data collection with different populations (i.e., are not specific to collecting data with young people).

1. Validity, reliability, and bias:

One of the most important concerns in collecting data on young people is that it be comparable over time. When data are comparable, researchers and programmers can identify trends and changes in the HIV epidemic among young people. To achieve this, data must be *valid*, *reliable* and *unbiased*.

There is no way to ensure that data are perfectly valid, but in some cases, invalid data can be clearly identified and rejected. Many errors can occur in data collection that can affect the validity of the data: there could be errors in the way the sample was selected, or with the way the interview was conducted, or even with how the data was analysed. One of the goals of this chapter will be to help readers ensure that they collect valid data.

Equally important is reliability. To assess reliability, researchers may repeat questions within an interview in exactly the same or somewhat different forms, or they may repeat them during a follow-up interview (Smith and Morrow, 1996). For example, if early on in a questionnaire, a respondent reports that he has never used a condom during sex, and then later, he reports that he used a condom the last time he had sex, his responses would be unreliable and invalid. While unreliable data indicate that data are invalid, reliable data are not necessarily valid. An example of this is when respondents report consistently and reliably that they have never had sexual intercourse, when in fact, they have already done so.

- **Valid** data are data that are as close to the truth as possible.
- **Reliability** is the consistency of an individual's responses to the same or similar questions.

When data errors are systematic or follow a regular pattern, they are called bias. No data are likely to be completely free of bias, but its extent must be recognized and minimized. There are several types of biases that can occur in data collection. One type of general bias, called **sample bias**, occurs when the respondents who participate in a study (the sample) do not adequately represent the group of interest. Having people volunteer to be in a study, or having public officials choose which villages or provinces are selected for a survey makes it very likely that the sample will be different from the rest of the group of interest. Sample bias is also likely to occur when people or

organizational units are selected for a survey but refuse to participate. Some researchers have identified this special type of sample bias as **participation bias**. An example of this is when there is a difference in past sexual risk behaviour between those agreeing to be tested for HIV and those refusing. To minimize sample bias, it's important to select random samples of participants, which are samples in which each member of the population has an equal chance of being selected. Details about random sampling are provided on pages 5-6.

A second general type of bias is called **information bias**, which occurs whenever the data show systematic differences across the sample. The two most common types of information bias are interviewer bias and recall bias. An interviewer, in particular, can influence the responses in many ways, even by the tone in his/her voice. Other characteristics, such as gender, age, race, ethnicity, education status, and attitudes may also influence how a study participant responds to questions. Researchers have found that participants are generally most likely to develop trust with people who are like themselves, and thus may report sensitive behaviour to such interviewers. For example, the most appropriate person to interview a rural adolescent girl may be a young woman of a similar background.

Another way interviewer bias can occur is when interviewers systematically code the data differently. For example, if respondents are asked to describe all family planning methods they have used in the past, and some report "wearing a magical charm," bias may occur if one interviewer systematically codes this as "local/traditional methods," while another codes it as "no method." Interviewer bias can be reduced by ensuring all interviewers are well trained and tested in interviewing skills, the research protocol and research ethics. Interviewer training should also involve a discussion of the importance of valid data, ways to reduce bias, and the research goals themselves to foster the interviewers' sense of ownership and personal commitment to those goals even in the absence of supervision (Bachrach and Newcomer, 2002).

Recall bias may also arise when individuals give false information about a past event. The ability to accurately recall events changes with the time elapsed since their occurrence and with their frequency. For example, data on condom use at last sex could suffer from recall bias: someone who had sex a long time ago may not accurately recall whether they used a condom or not. A respondent who had sex very recently, however, could say for certain whether they did. To reduce recall bias, one simple technique is to ensure that participants have enough time to reflect before answering, and another is to have participants think through a sequence of events in their life history (Auriat, 1993; Bradburn et al., 1987).

A third type of bias that may be especially common among young respondents is social desirability bias, which occurs when study participants do not answer questions honestly because they perceive the truth to be socially unacceptable or undesirable. For example, a person may falsely deny having had a sexually transmitted infection because of fear of the social stigma related to it. Researchers have generally found that behaviours that are seen as socially undesirable will be under-reported rather than over-reported (Catania, 1990). However, perceptions of desirability may vary within a population, and what may be considered unacceptable for one group may be considered desirable for another. In some cultures, for example, young women's abstinence from sex is highly valued, so they may under-report their total number of sexual partners in research studies, while

young men in the same culture may take pride in sexual experience, and thus exaggerate their total number of partners (Siegel et al., 1998; Catania, 1993).

A number of techniques have been developed to reduce social desirability bias. One is to explain the research goals before the interview, including their social value and legitimacy, and the importance of honest reports. Another is to ensure the privacy of the interview and the confidentiality of any information obtained from it, and to make this as evident as possible to respondents.

- **Interviewer bias** occurs when there are differences in how the interviews are being conducted and the data being coded.
- **Recall bias** happens when people systematically make errors when remembering events.
- **Social desirability bias** occurs when study participants do not answer questions honestly because they perceive the truth to be socially unacceptable or undesirable

2. Questionnaire Design and Wording

Questionnaire design and wording may also contribute to bias. For example, if respondents are asked about sexual behaviour early in the questionnaire, social desirability may be greater than if they were asked later, after becoming more familiar with and trusting of the interview process. The length of the questionnaire and the duration of the interview may also be important, as if it takes too long respondents may become tired or frustrated, and stop responding in a meaningful way (Smith and Morrow, 1996).

There are other ways in which the order and content of questions and answers can influence respondents' reports. If the wording of the question suggests a possible answer, respondents may consider it and report it when they would not have otherwise. Similarly, early questions and answers may suggest and influence answers to later questions (Smith and Morrow, 1996). As an example, if the question, "Have you ever used condoms to prevent pregnancy?" is asked before the request, "Name all family planning methods that you know," the first question will have provided one possible answer to the second that some respondents may not otherwise have considered or reported.

Another problem often contributing to misunderstanding on the part of the researcher or the respondent has to do with vague or relative terminology, such as "often" or "frequent" (Bradburn and Miles, 1989). For example, if respondents are asked how often they have had sex, one young man may feel he has sex often because he has sex once a month, whereas another may feel he does not have sex often, because he only has sex once per week. Ideally, questionnaires should avoid such vague terms, or if necessary, clearly define them.

In many developing countries, the first languages of researchers and respondents may be different, and questionnaire translation becomes a potential source of error or bias (Smith and Morrow, 1996; Bulmer and Warwick, 1983). Error may simply occur due to the poor quality of the translation, but also because many terms may not have conceptual equivalents in the other research language(s), and/or may have multiple, undesired meanings. For example, in English the term “to seduce” suggests successfully persuading someone to have sex, but in another language the closest conceptual equivalent may only suggest *an attempt* at seduction (not necessarily resulting in sexual intercourse). If the two terms are used interchangeably, reports of attempted seduction may be misunderstood as seduction that results in intercourse.

The issue of mis-translation is particularly problematic in sexual behaviour research, because its sensitive nature makes it difficult to identify appropriate terms. In many cultures, for example, the literal translations for “vaginal intercourse” are obscure or are considered to be offensive. In contrast, euphemisms such as “to make love” may be widespread and relatively inoffensive, but also may be ambiguous and suggest other activities to respondents, such as sexual play/exploration, non-penetrative sex, or anal intercourse.

During questionnaire development, there are a number of ways to reduce the possibility of bias. If a draft questionnaire requires translation, a new translator should do a **back-translation** into the original language, for researchers to be able to assess errors and unanticipated meanings. This process can be repeated several times during questionnaire development to ensure the best quality of translation. **Pre-tests** can be done with a small number of respondents, focusing on specific questions and following up closely to see if there were any areas of misunderstandings. Pre-tests are particularly useful for clarifying problem issues, such as appropriate sexual terminology to use in research with young people. Once the study protocol has been drafted, **pilot tests** (small-scale models of the study itself) can take place in a similar setting to that proposed for the study and with respondents who are similar to the anticipated study population (Smith and Morrow, 1996).

Tips for designing the questionnaire:

- Begin with general, non-sensitive questions (i.e., avoid asking whether a young person has ever had sex in the first few questions)
- Use simple formatting and wording.
- Use language familiar in the culture and sub-group.
- Avoid ambiguous terms or, if necessary, clearly define them (i.e., for some populations, it may be important to define the meaning of sex before it is asked in questions).
- Avoid a question that suggests a response to that question or to a later question

Tips for developing and testing the questionnaire:

- Arrange for a series of independent translations and back-translations. The questionnaires need to be translated into the respondents’ local language before the survey begins. In a separate operation, another translator should translate the new questions back into English (or the original language) without referring to the original questionnaire. This new translation should match the original version.
- Pre-test questionnaire to identify problem areas, misinterpretations, or cultural objections to the questions

3. Data Collection Methods

A range of research methods have been used for adolescent sexual behaviour research, each of which has strengths and weaknesses in reducing biases and increasing data validity. **Quantitative** methods involve collecting data from a large number of people for statistical analysis. For practical reasons, the large scale of quantitative studies usually requires that surveys be **structured**, or follow an exact question and answer format. The uniform nature of such structured questions can help minimize interviewer bias. For example, if a structured interview is conducted correctly, there is less chance of interviewers using different terms for the same questions, and thus less chance of suggesting a different possible answer to respondents. **Qualitative** methods generally involve collecting more in-depth information from far fewer people than in a quantitative study, using semi-structured and unstructured interviews. Qualitative interviews allow for a more detailed and complex exploration of topics than is usually possible in quantitative interviews, and since they are less structured, researchers can adapt and respond to issues that arise spontaneously. For example, a study focusing on sexual behaviour may have no specific questions about violence in relationships, but if a respondent unexpectedly reported engaging in high-risk behaviours due to a threat of violence, there would be scope to explore this in a semi-structured interview.

The recommended methodology of data collection throughout most of this guide is a national or regional **household survey** of young people. **School-based surveys** are also used occasionally to collect data on young people. One of the main advantages of school surveys as compared to household surveys is the cost: school surveys are, overall, much less expensive to implement. The lower cost of these surveys is largely due to a more accessible sample (the advantage of having a group of young people in a school, rather than having to find each young person in his/her household), and the fact that most school surveys use self-administered questionnaires (that is, they do not require an interviewer to pose questions, as young people fill out the questionnaire themselves).

Another difference between these methodologies is that school-based surveys, in comparison with household-based surveys, tend to produce a higher prevalence of adolescent risk behaviours (see Brener et al 2003; Kann et al, 2002). The main hypothesis for this finding is that privacy appears to be the critical determinant of whether young people are willing to report risky (often stigmatised) behaviours (Kann et al, 2002). This is in fact a disadvantage of household surveys: they are often unable to ensure the privacy for respondents and this might impact the report of risk behaviours. However, evidence to establish if the higher prevalence found in school-based surveys is closer to "true" prevalence has not been conclusive, and there is also evidence to suggest that data from school surveys may be less valid than data from household surveys (see Gregson, Simon: "The performance of face to face interviews and

self-administered/more anonymous techniques," Workshop on Measurement of Sexual Behaviour in the Era of HIV/AIDS. London, 4th to 6th September 2003.)

The major drawback of school-based surveys - and the main reason why they are not suggested as the means of data collection in this guide - is that they are not representative of the overall population of young people. There are two main arguments:

- This guide recommends most of the indicators to be collected with young people aged up to 24 years. In countries where school attendance is high, young people in the higher age bracket (18 and above) are no longer in school, and would therefore be excluded from the sample.
- In countries where overall school attendance is low (and usually lower with increasing age), a substantial proportion of the overall young population would not be included in a school-based sample.

Indicators in this guide were compiled for the purpose of monitoring and evaluating national programmes, and for tracking national behavioural and biological trends relevant to HIV. Given the limitations specified above, school-based surveys cannot be substituted for household surveys.

Nevertheless, school-based surveys can be a valuable additional source of information, which should be used in conjunction with a household survey to give insight into specific issues, particularly those related to school-based interventions. There are a number of high quality school-based surveys which have been applied in many countries:

- The Global Youth Tobacco Survey (sponsored by the World Health Organization and Centers for Disease Control), is survey of tobacco use and its determinants among 13-15 year old students. It has been conducted in 140 countries.
- The Health Behaviour of School-going Children Survey (sponsored by the World Health Organization), is a survey of health behaviours and their determinants among 11, 13 and 15 year old students. It has been conducted in 29 countries.
- The Global School-based Student Health Survey (sponsored by the World Health Organization in collaboration with UNICEF, UNESCO, UNAIDS and CDC), is a survey of priority health risk behaviours and their determinants among 13-15 year old students. It was launched in 2003 and has, to date, been conducted in 7 countries.

Ultimately, the most important point is the synergy of different methods and efforts of data collection in providing a comprehensive picture of issues relevant to preventing and treating HIV/AIDS among young people. All surveys - whether household-based, school-based, or with special populations - should include a key set of core indicators relevant to HIV and young people. The indicators in this guide are such a proposed set, with the aim contributing to comparability of data across surveys, across regions and across time, and therefore improving data collection efforts overall.

For most of the indicators included in this guide, data will need to be collected by conducting a national or regional **household survey** of young people. Household surveys are widely used to collect population-based national- and sub-national-level data on the health status and behaviours of adolescents and young adults. Among the more prominent ongoing household surveys that collect data in a sizeable number of countries include the Demographic and Health Surveys (DHS), the Behavioural Surveillance Surveys (BSS) developed by Family Health International, and the UNICEF Multiple Indicator Cluster Surveys (MICS). These surveys are based on adult samples, and the sample size and sampling techniques have to be adjusted to ensure a representative number of young people within the overall sample. There are also household surveys that are adolescent-specific, such as Reproductive Health and Adolescent Reproductive Health Surveys undertaken by the U.S. Centers for Disease Control and Prevention (CDC). In addition to these “international” surveys, many countries also undertake annual or more frequent national surveys (e.g., quarterly labour force or economic surveys). The volume and types of data on adolescents varies tremendously across these national surveys. However, as such surveys might serve as a vehicle for adding questions on adolescent health and development, they are worthy of note.

Advantages and Disadvantages of Household Surveys

Advantages:

- High feasibility of collecting national representative data on youth, both in-school and out-of-school
- Lower susceptibility to sample bias
- Considerable experience with acceptance of surveys as a means of monitoring by most governments around the world
- The relative ease of standardization of surveys across settings, which makes intra-national and international comparison more feasible.

Disadvantages:

- Moderate-to-high cost
- Relative inefficiency in obtaining measurements for vulnerable sub-population of youth (i.e., street children, orphans who do not have a fixed place of residence), the result of a lack of suitable sampling frames
- The fact that youth are highly mobile and often not at home may make it difficult to find them to be interviewed.
- Difficulty in ensuring privacy and the subsequent impact this may have on behavioural prevalence estimates

For an in-depth discussion of different surveys and methods of sampling populations of young people, refer to the "Guidelines of sampling of youth" paper, which is available on www.childinfo.org.

4. Sampling

Sampling is the process of systematically choosing a sub-set of the total population to be studied. Due to their lower risk of bias, surveys undertaken using **probability sampling methods** are generally preferred to those based on **non-probability samples**. For collecting data on the indicators of this guide, it is critical that probability sampling methods be used as they are the only way that the data can be compared over time. Probability sampling methods require that every person in the population has a chance of being selected for the survey. All of the international household surveys mentioned previously use some form of probability sampling, typically multi-stage cluster sampling. Conversely, non-probability sampling methods are not based on probability theory. With these methods, samples are not chosen by random chance. Examples of non-probability sampling methods can include obtaining a sample of subjects on the basis of opportunity (i.e., surveying all youth you meet walking down a street), or basing sample selection on referrals from other sample subjects (often referred to as snowball sampling).

There are several types of probability sampling methods. The simplest of these methods is **simple random sampling**, in which subjects are chosen at random so that each element has an equal chance of selection. For example, survey subjects are chosen from a hat or, ideally, from a table of random numbers in a statistical textbook. Another type of probability sampling method is **stratified sampling**, in which the population to be sampled is divided into homogenous groups based on characteristics that are important to the indicators being measured, such as youth who are sexually active. A simple random sample is then chosen from each group. Large-scale surveys, like DHS and MICS, are based on **multi-stage cluster sampling**, which typically involves selecting random samples of clusters (such as districts or regions in a country) and then enumerating all the households within these districts, and taking a simple random sample of each cluster.

The availability of a sampling frame is an important requirement for probability sampling. Indeed, it is the lack of a usable sampling frame and the high cost of creating one from scratch that is often the rationale for using non-probability methods. For the general population of adolescents, usable sampling frames in the form of a recent census, “master” sampling frames developed by the national statistics office, and/or other sampling frames used by other large-scale surveys are likely to be available in most countries. However, usable sampling frames for vulnerable sub-groups of youth (i.e., commercial sex workers) are not likely to exist in most settings, and either appropriate sampling frames will have to be developed or non-probability methods resorted to.

Probability sampling methods are based on **probability theory**, a mathematical concept based on accepted statistical principles that refers to the ability to predict the statistical likelihood that a random event will occur.