Collecting and Analyzing Data
Data Collection

The following chapter is excerpted from *Designing HIV/AIDS Intervention Studies: An Operations Research Handbook*, Andrew Fisher and James Foreit, 2002, Washington, DC: Population Council. (More on OR Handbook)

CHAPTER 9

DATA COLLECTION

Discussion of how data will be collected is an important part of the methodological section of your proposal. There are many different ways to collect data. The approach you choose depends on the study objectives, the study design, and the availability of time, money, and personnel. In deciding on the best way to collect data, it is important to consider whether the study is intended to produce **quantitative**, numerical findings or to produce **qualitative**, descriptive information.

Most operations research studies are concerned with the quantitative measurement of program operations, but many also are (or should be) concerned with detailed qualitative information on processes (for example, how a project is actually implemented in the field, how couples decide to use condoms, or how PLHA can become more involved in the implementation of programs). Often, study objectives call for both quantitative and qualitative information, which may require that you use more than one data collection method.

Quantitative Data

One of the most common ways to collect quantitative data on people is to use a standard questionnaire that is administered by a trained interviewer. There are other ways to collect quantitative data, including self-administered questionnaires, service statistics, or such secondary sources as the census, vital records, an HIV/AIDS sentinel surveillance system, or other existing records and reports.

If a study's sample is composed of geographic or organizational units (such as villages, districts, clinics, hospitals, and VCT centers) rather than people, quantitative data usually can be obtained from service statistics and secondary sources. If information is not available from such sources, it may be necessary to obtain data on geographic or organizational units by interviewing people who are members of these units or are knowledgeable about them.

Structured Interviews

Studies that obtain data by interviewing people are called **surveys**. If the people interviewed are a representative sample of a larger population, such studies are called **sample surveys.** If the sample is large enough to permit statistical analysis, it is customary to use structured interviews rather than unstructured ones, since the former lend themselves better to quantitative analysis and the latter create serious data processing difficulties, particularly if the sample is large.

A structured interview is one that uses a standard questionnaire (or interview schedule) to ensure that all respondents are asked exactly the same set of questions in the same sequence. The exact wording of each question is specified in advance, and the interviewer merely reads each question to the respondent. In designing a questionnaire and then using interviewers to administer it, you need to remember several points:

Use simple language that can be easily understood by the respondents.

Precode the responses to the questions whenever possible so that the information can be transferred easily to a computer and then tabulated. This requires more effort when designing the interview schedule, but the time saved during the processing and analysis more than compensate for it.

Try to avoid embarrassing or painful questions. If it is necessary to ask a sensitive question, word it as tactfully as possible and avoid asking it near the beginning of the interview, when the respondent is less relaxed. It is generally best to put sensitive questions in the middle of the questionnaire.

Don't use leading questions that strongly suggest a particular response. For example, in a survey of sex workers, the following would be a leading question: "Most sex workers in the world experience violence from their clients. Have you also experienced violence from your clients?"

Avoid asking for more than one item of information in a single question. For instance, do not ask, "Do you and your husband want another child?" If the respondent and her husband disagree about having another child, an answer of either "yes" or "no" will be impossible to interpret accurately. A response of "yes" could mean "I want another child," "My husband wants a child," or "Both of us want a child." It would be better to ask two separate questions: "Do you want another child?" "Does your husband want another child?"

Watch out for ambiguous wording of questions. For instance, if you are conducting a survey among women, a question such as "Do you use a contraceptive when you have sex with your partner?" may seem clear enough. But since the respondents are women, some of them may answer "no" even though their partner has had a vasectomy or uses a condom; these are male methods used by men, not women. A better way to word the question would be to ask, "Do you or your partner use a contraceptive when you have sex?"

Do not overload your interview schedule with questions that are not essential for your study. Keep it as short as possible to avoid tiring your respondent and to simplify the data processing and analysis.

Include all questions necessary to provide sufficient information on the variables you want to study. Also be sure that the data necessary to test the hypotheses of the study can be obtained from the questionnaire instruments. It is often helpful to prepare a list of key study variables with an indication of where the data for each variable will be obtained. For example, to be sure that all variables in your study have a source of information, you should create the following type of table:

Variable	Source of Data
Age	Question 5
Attitude toward HIV/AIDS	Questions 8, 10, 26, 28, 32
HIV status	Sero laboratory test
Contact with peer educator	Questions 14 and15
HIV prevalence district	Sentinel surveillance

Start with the easier questions, and move on to the ones that are more sensitive or difficult after the respondent has had an opportunity to become accustomed to the interview situation. Respondents are likely to be somewhat tense or even suspicious at the beginning of the interview, so a major task of the interviewer during the first few minutes is to establish rapport in order to put the respondent at ease. This task is easier if the initial questions are not likely to cause embarrassment or be difficult to answer.

Ask all respondents each question in exactly the same way. If the interview is to be conducted in more than one language, prepare full written translations in all major languages and instruct your interviewers to use those translations word for word. Do not permit free translations, except for languages with too few respondents to justify the cost of preparing written translations. To ensure comparability of wording among the various written translations, have them "**back-translated**" into the original language to verify that the meaning is retained. The back translation should be done by persons who are not familiar with the original wording of the questionnaire.

Pretest the questionnaire in an actual field situation. Here are several principles of pretesting that should be noted:

1. The pretest does not need to involve large numbers of respondents; 30–50 respondents are often enough if they are sampled (purposively) in a way that ensures that the expected heterogeneity of the study sample is reflected in the pretest sample. This means making sure that the pretest includes the same types of respondents who will be included in the study sample: old and young, urban and rural, less educated and more educated, males and females, and so on.

2. Be prepared to conduct more than one pretest. If a pretest results in major revisions, it is a good idea to conduct a second pretest to be sure the revisions are satisfactory.

3. Pretesting should be completed before the training of interviewers. Often it is possible to use field supervisors to do the pretest. This gives them a clearer understanding of study objectives and better prepares them to help train the interviewers.

4. The main purpose of the pretest is to ensure that the respondents are able to understand the questions and answer them usefully. Hence, it is not enough simply to interview the pretest respondents; rather, each interview should be followed by a **debriefing**. During the debriefing the interviewer asks about the respondent's understanding of questions that are likely to be misunderstood or that appear to have caused difficulty during the interview.

Provide complete training for all interviewers. The training should be designed to familiarize the interviewers with the intent and meaning of the questions, let them role-play interview situations, and give them experience in actually conducting interviews in the field under supervision. Provide them with an instruction manual that clearly explains procedures for completing questionnaires. Be sure the training is of sufficient duration for the trainees to become skilled at interviewing. One of the most serious mistakes that can be made by a study investigator is to provide inadequate training to interviewers.

An appointment should be made for an interview callback if a prospective respondent is not available during the interviewer's first visit. It is common to require at least two callbacks before dismissing a sample case as unavailable for interview.

Interviewers must be given clear instructions for obtaining substitutes. If a study sample is small, it may be necessary to find substitutes for cases that cannot be located. Interviewers need to know how to select substitute cases that will ensure either a random sampling of substitutes or a selection of substitutes who are similar to the cases originally selected. However, if the sample is large enough to tolerate some loss of cases, it is usually better not to use substitutes.

Isolate the respondent during the interview. If other people are present, the respondent's answers may be influenced by them. For example, if a man is interviewed in the presence of his wife, he may not be entirely honest in replying to a question about the number of casual partners he has had sex with in the last month.

Check all completed interview schedules for errors, omissions, and discrepancies as soon after interviewing as possible. Respondents should be revisited to correct errors that cannot be otherwise resolved. It is best to have the interviewer check the questionnaire immediately after the interview so that the respondent can be consulted. After the interviewer has checked and corrected the questionnaire, it should be rechecked by the field supervisor. This checking process is known as **field editing**.

Service Statistics

All national HIV/AIDS control and coordinating organizations generate program statistics, as do many service delivery organizations. Some organizations have established a management information system (MIS). The quality of service statistics, however, varies from country to country and even within countries and thus these statistics should be used with caution.

Service statistics often help researchers define the parameters of the problem they want to study. In some cases, service statistics can be used to compare the results of a particular study with nationwide figures. In operations research projects it is often necessary to design supplementary forms to provide data that are not available from the regular service statistics.

Service statistics have the advantage of being an inexpensive data source. Using service statistics instead of surveys can save many thousands of dollars in research costs. However, the type of data available is more limited than data obtained from a survey, and service statistics data often have serious reliability problems. Two common problems are that they are often incomplete and that providers are not trained to correctly fill in the forms. Before doing a study that relies on service statistics, be sure to assess the reliability of the data.

Self-administered Questionnaires

If money, personnel, and time allow, interviews are generally preferable to self-administered questionnaires. For most surveys in developing countries, self-administered questionnaires are difficult if not impossible to use because many respondents are not educated enough to complete questionnaires themselves.

There are other problems with self-administered questionnaires, even among educated respondents:

- Instructions or questions are more likely to be misunderstood without an interviewer to help explain them.
- Portions of the questionnaire are more likely to be left blank.
- It is difficult to incorporate many conditional sequences of questions (for example, "If the response to question 12 is 'yes,' go to question 13; if not, skip to question 18") without causing confusion.

Self-administered questionnaires are likely to be useful in situations where literate respondents are already gathered together in a setting where they can, for example, write in a classroom or an office. Self-administered questionnaires can be especially useful in evaluating school-based HIV/AIDS programs or training programs for peer educators. They may also produce more accurate results on such sensitive topics as sexual behavior.

Sometimes, blank questionnaires are mailed out to respondents, who are asked to complete them and send them back. This method has the virtue of being very inexpensive but also has all the drawbacks noted above, plus the added problem of high nonresponse rates. It is common for mailed questionnaires to elicit less than a 10 to 20 percent return, even after one or two reminders. This drawback affects how representative the sample is and may render the validity of quantitative findings so questionable that they are of very little use.

Situation Analysis

An approach that has been used in recent years to collect information on the functioning of an entire service delivery system is called a **Situation Analysis**. This approach can be used to examine, for example, how prepared the formal health care system of a country is to address the needs of HIV/AIDS clients.

The Situation Analysis approach involves visiting a relatively large number of randomly selected service delivery points (SDPs), usually clinics or hospitals; interviewing managers, providers, and clients at these SDPs, using structured interviews; conducting a full inventory of the SDP's equipment and commodities on the day of the visit; and observing the client-provider interaction at the SDPs on the day of the visit. This methodology allows you to collect a relatively large and detailed amount of information on the functioning of the health care system and the type and quality of HIV/AIDS services that are available. A limiting factor is that when a large number of SDPs are sampled, the cost of conducting a Situation Analysis study can be high.

Secondary Data Sources

Information from recent censuses, vital statistics, sentinel surveillance systems, UNAIDS, WHO, ILO, the World Bank, and even previous surveys can often be used with data collected especially for a study to enrich the analysis. A large body of data on health and HIV/AIDS has been collected by many organizations and by national surveys such as the Demographic and Health Surveys. Much of this information is available through the internet (see the list of internet addresses in the bibliography).

Content Analysis of Written Materials

This method is usually not used as the sole method in an HIV/AIDS operations research study, but it can serve as a useful adjunct to other types of data collection and analysis. For instance, the content of documents related to a training curriculum may be analyzed to determine what type of knowledge and skills the training is supposed to develop. The content analysis can guide the researcher in devising procedures to test the knowledge and skills of the trainees.

Information and education materials on HIV/AIDS can be "content analyzed" to indicate whether messages are being overemphasized or underemphasized. The content of press reports or public statements made by policymakers can be studied to assess attitudes toward HIV/AIDS and issues concerning HIV/AIDS stigma and discrimination. Research reports may also be content analyzed to determine the current state of knowledge about a particular research topic.

Qualitative Data

The data collection techniques most appropriate for studies whose objectives call for descriptive, qualitative data tend to be different from those most appropriate for quantitative analysis. Operations research studies often use a combination of quantitative and qualitative data collection methods to obtain the most accurate and realistic picture of a program situation. Quantitative methods discussed earlier are important for obtaining data for making predictions, "probabilistic" statements, and generalizations. Qualitative methods such as unstructured interviews, focus group discussions, and direct observation of operations are important to obtain data on processes, on how and why a program works, and on unintended and unanticipated program outcomes.

Unstructured Interviews

The chief drawback of structured interviews is that the responses obtained tend to be short and sometimes superficial. An alternative approach to interviewing, which permits greater depth of meaning, is to seek detailed, open-ended responses to questions. Such interviews are often called **in-depth interviews**. Instead of reading formal questions from a structured interview schedule, the interviewer has an outline of topics or a set of general questions to serve as a guide to the kind of information required. Details that are not brought out initially are sought through follow-up questions called **probes**.

The chief drawbacks of unstructured interviews are that (1) the interviews require highly skilled and experienced interviewers, and (2) the analysis can be complex and time-consuming. The shortage of qualified interviewers and analysts and the high cost of conducting and processing such interviews usually mean that a small sample size must be used (sometimes as few as 20–30 respondents).

In-depth interviews are usually most useful in exploratory studies that seek to clarify concepts or generate hypotheses before developing questionnaires for quantitative surveys. They also are useful for generating supplementary, explanatory data to augment the findings from larger surveys. For example, little is known about the gender and power dynamics involved in the negotiation between sex workers and their clients regarding the use of condoms. An unstructured interview could help gain insight into the dynamics and determinants of the decision to use or not to use condoms.

Focus Group Discussion

A way of reducing the amount of time and number of personnel required for conducting and analyzing in-depth interviews is to bring respondents together in discussion groups that focus on a particular topic. The use of **focus group** discussions

has the advantage of being economical yet still yields detailed qualitative information from a relatively large number of respondents. It is often an excellent technique to use for examining group or community consensus about a particular issue.

The interviewer (or facilitator) follows nearly the same procedure as in unstructured interviews, using a general discussion guide and eliciting details through probes. Participants are usually sampled purposively to reflect population variations that are particularly relevant to the research topic. For instance, cases sampled for focus group discussions might consist of a group of truck drivers, adolescents, sex workers, teachers, or another group of interest. While focus group discussions can generate extremely valuable information, they are not easy to conduct. To obtain meaningful information, a highly skilled and trained facilitator must guide the group but not lead it in a predetermined direction. Also, the analysis of transcripts from focus group discussions is not easy; if not done correctly, this can sometimes generate findings that are more fiction than fact.

Direct Observation of Operations

This technique can generate either quantitative or qualitative data, but tends to be used more for small-scale exploratory studies than for large-scale quantitative studies. The reason for this is that it usually requires relatively highly skilled observers and analysts and prolonged periods of observation, resulting in a high cost per unit of observation. This technique lends itself particularly well to observation of community responses to program efforts. It is the chief method of ethnographers, who specialize in community studies. It is also useful for organizational studies, such as observation of clinic operations, activities of field workers, and administrative procedures. The researcher should note, however, that when workers are being observed they often do not behave in ways that are typical of their day-to-day behavior.

What To Do: Data Collection

1. Review your study objectives and hypotheses as well as your list of independent and dependent variables. What types of information do you require? Which data gathering technique(s) would be most appropriate and feasible for gathering the desired information? Is some of the information already available from other sources?

2. If you intend to collect information through a survey, review the list of points under the heading Structured Interviews. Be sure your proposal discusses important steps, such as translation, pretesting, training of interviewers, and rules about callbacks and substitutions.

3. Make an outline of the data gathering instrument(s) you intend to use (for example, the interview schedule, discussion guide, or observation guide). Give examples in the proposal of questions to be asked, especially those designed to elicit information about key variables found in the hypotheses. Be sure you incorporate measures (or at least descriptions of measures) of all the variables you intend to study. It is useful to list all the variables of the study and then under each variable record the question or questions that will be used in the questionnaire. Here are some examples:

   Variable 1: Condom use
   Q1: Have you ever used a condom?
   Q2: Did you use a condom the last time you had sex?
   Q3: Do you use a condom with casual partners?

   Variable 2: Education
   Q4: What is the highest grade in school you have completed?

4. Describe your data collection procedures and include the description with the outline of the data gathering instrument(s) in the proposal.

Data Quality Checks

There are several ways to check the quality of interview data:

- Sometimes a researcher will deliberately ask two or more questions that yield the same type of information. The first question might be asked at the beginning of the interview and the second at the end. The two questions are then examined for consistency of response. This is one way to check the reliability of the data.
- For difficult questions, sensitive questions, or questions for which the researchers want to be sure the information is correct, the interviewers can be instructed to probe. That is, the interviewer can repeat the question in a slightly different form or repeat the respondent's answer and then ask if the information is correct. For example, a woman might report that she has two sons and three daughters. The interviewer might then say, "You have a total of five children, two males and three females. Is that correct? Are there any other children you may have forgotten to tell me about?"
- Field supervisors should be used to help the interviewers with difficult situations and to make sure that they are actually doing their work. (Occasionally, interviewers complete their questionnaires in tea stalls or beer halls!) Some studies use a ratio of one supervisor for every five interviewers.
- For most studies using an interview procedure, an attempt is made to re-interview a certain percentage of the respondents. Depending on the size of the sample, a general rule is to re-interview between 5 and 10 percent of the sample. The data from the first interview are then checked against the data from the second interview for consistency. This is another check on the reliability of the data. Obviously, if there are major inconsistencies, particularly on such demographic profile questions as age, marital status, and parity, there is a problem somewhere. The problem might be with the questionnaire, the interviewers, the tabulation procedures, or something else.
- Once the data have been collected and tabulated, it is possible to do statistical checks for errors or for consistency of response. For example, a frequency distribution of the parity of women may reveal that several women claim to have 18 or 19 living children. Since this is highly unlikely, the investigator is faced with the choice of either discarding the entire questionnaire, eliminating at least the information from the question on parity, or going back and re-interviewing the women who claim to have 18 or 19 children.

What To Do: Data Quality Checks

Describe the procedures you will use to check the quality of the data collected. Consider the following procedures:

1. Include repeat questions in your questionnaire that can be used to check for consistency of response.

2. Have supervisors monitor the work of the interviewers in the field.

3. Re-interview a percentage of respondents and look for inconsistencies.

4. Recode a percentage of the questionnaires to be sure that there are no coding errors.

5. Examine the frequency distribution on all variables to see if there are odd codes or items that are not logical.

Confidentiality of Information

It is always important to maintain the confidentiality of the information collected from respondents. Unless absolutely required, do not collect information that is sensitive or potentially harmful. Whenever possible, use code numbers instead of names. Assure the respondents that the information they give will be kept confidential. Do not let other people use the information you have collected when there is a chance that the use of the information could be damaging to the respondent. You have an obligation to protect the confidentiality of the respondents in your study.

Operations research on the topic of HIV/AIDS often involves asking very sensitive questions and collecting information that, if disclosed, might create problems for the respondent. Questionnaires should always be stored in a closed, locked cabinet. During the training of interviewers and supervisors, it is extremely important to thoroughly cover the topic of confidentiality and informed consent. If respondents do not want to be interviewed, you have an obligation to respect their wishes. Use an **informed consent form** to explain the basic nature of the study and obtain the agreement of the respondent to be interviewed. The informed consent form should be written in simple, plain language that is understandable to everyone. During data collection, it is important to make spot checks to be sure that the informed consent form is being used by the interviewers.

What To Do: Confidentiality of Information

1. Describe in detail how you plan to maintain the confidentiality of information collected, including how you are going to store questionnaires and use code numbers instead of names.

2. Describe the training process you will use to instruct interviewers and supervisors on the use of the informed consent form.

3. Include a sample of the informed consent form in your study proposal.