bout the Book

Data Collection, Management and Analysis in Academic Research: Proceedings of a workshop is a collection of papers esented at a workshop organized by the Post- ...chool of the University of Ibadan.

t explores various topics which include data nalysis, management, collation and collection of data and use and the importance of data in research. It is recommended to all scholars and researchers.

978-978-488565-0-8

DATA COLLECTION, MANAGEMENT AND ANALYSIS IN ACADEMIC RESEARCH

Edited by: Labode Popoola, Olajide Olorunnisola and Olusegun Ademowo

DATA COLLECTION, MANAGEMENT AND ANALYSIS IN ACADEMIC RESEARCH

Proceedings of a Workshop

Edited by

Labode Popoola

Olajide Olorunnisola

Olusegun Ademowo

Published by : The Postgraduate School
University of Ibadan,
Ibadan

E-mail: postgrad@mail.ui.edu.ng

Website: http://ww.postgraduateschool.ui.edu.ng

First Published 2009

ISBN 978-978-48855-0-8

Printed in Ibadan by Sapphire Prints
08034892535

CONTENTS

CHAPTER 3

DATA COLLECTION TECHNIQUES

O.I. Ajewole and A. Odaibo

Introduction

Data collection is a crucial stage in the planning and implementation of a study. If the data collection has been superficial, biased or incomplete, data analysis becomes difficult, and the research report will be of poor quality.

Data collection techniques are designed procedures to systematically collect information about our objects of study (people, objects, phenomena) and about the settings in which they occur. Collection of data has to be systematic. If data are collected haphazardly, it will be difficult to answer research questions in a conclusive way. The two basic methods for data collection are through :

- a controlled experiment (Experimental studies); and
- a survey (observational studies).

Experimental Studies

An experiment is a structured and organized method used in determining the relationship between the different factors (Xs) affecting a process and the output of that process (Y). It is an investigative approach where the system under study is under the control of the investigator.

This means that the individuals or materials investigated, the nature of the treatments or manipulations under study and the measurement procedures are all determined, at least in their major and important features, by the investigator. Experimental studies that involve humans are called clinical trials and their

purpose is usually to draw conclusions about a particular procedure or treatment. In research, the word control is used in two ways: to reduce or eliminate outside influence on variables and to denote a control group for comparison. In experimentation, all relevant factors are varied systematically. When the results of these experiments are analyzed, they help to identify optimal conditions, the factors that most influence the results, and those that do not, as well as details such as the existence of interactions and synergies between factors. Examples of experiments include investigations carried out in Laboratories, Green houses, and Fields under specified conditions.

Measurement of Experimentation Data

Three questions are to be considered when measurement issues are discussed: which measurements will best suit the research objectives; what additional measurement could be useful and at what scale; when will the measurements be taken; and how will the measurements be made and recorded.

When planning measurements to be taken, it is important to clarify whether particular measurements will be taken at the experiment level, or at the plot (unit) level, or at the plant (sub-unit) level. For example soil measurements may be taken at the experiment level to characterise the site, or within each plot to include within the formal analysis.

Occasionally there are unplanned events that necessitate a review of the objectives of the trial. For example, if the level of rainfall is low, in a variety trial, should irrigation be applied? This would keep the stated objectives, but make it more difficult to specify the domain of recommendation for the experimental site. An alternative would be to reconsider the objectives and perhaps to study drought tolerance instead of yields. The new objective(s) would lead to different measurements from those planned earlier. For any experiment, measurements to be taken can be considered in three phases.

Measurements taken before the experiment begins
These can be for two purposes. The first is to assess which site is appropriate for the trial and will often involve a small survey. The second purpose is to record the initial conditions before the trial begins, for example, soil constituents, animal weights or initial plant stand.

Measurements taken during the course of the experiment
Possible measurements include labour use for different operations, weed weights, tree height, dry matter, quantity of food eaten by animals, animal weights, disease incidence, etc.

Measurements taken at the end of the experiment e.g. yields in crop trials, degree of damage in storage trials, wet/dry weight of fish in aquaculture experiments, germination rates in seed storage experiments or milk yields in comparison of diets in a livestock trial.

Data Collection through Sample Survey (Observational Studies)
A survey is a research process in which new information is collected from a sample drawn from a population, with the purpose of making inferences about the population in as objective a way as possible. The basic idea behind survey methodology is to measure variables by asking people questions and then to examine relationships among the variables. In most instances, surveys attempt to capture attitude or patterns of past behaviour.

The term "survey" actually refers to one, or some combination of two, procedure(s): questionnaires; and interviews. A questionnaire almost always is self-administered, allowing respondents to fill them out themselves. All the researcher has to do is arrange delivery and collection. An interview typically occurs whenever a researcher and respondent are face-to-face or communicating via some technology like telephone or computer. There are three subtypes of interviews: unstructured, which allows spontaneous

communication in the course of the interview or questionnaire administration; structured, where the researcher is highly restricted on what can be said; and semi-structured, which restricts certain kinds of communication but allows freedom on discussion of certain topics.

Various survey techniques include:

- Using available information;
- Observing;
- Interviewing;
- Administering written questionnaires;
- Focus group discussions; and
- Projective techniques, mapping, scaling.

The aforementioned data collection techniques can be broadly categorized as qualitative or quantitative research techniques. Qualitative research techniques involve the identification and exploration of a number of often mutually related variables that give INSIGHT into human behaviour (motivations, opinions, attitudes), in the nature and causes of certain problems and in the consequences of the problems for those affected. 'Why', 'What' and 'How' are important questions. They produce qualitative data that are often recorded in narrative form. Qualitative research techniques include flexible techniques, such as loosely structured interviews using open-ended questions, focus group discussions, and participant observation.

Quantitative research techniques on the other hand are used to quantify the effects of systematic variation of one or more independent variables on some other response or dependent variables; or the size, distribution, and association of certain variables in a study population. Thus, 'How many?' 'How often?' and 'How significant?' are important questions in quantitative research techniques. Experiments and structured questionnaires are good examples of quantitative research techniques. The answers to questions can be counted and expressed numerically while variations of processes can be measured by interval scale instruments.

1. Using available information or secondary data

Usually there is a large amount of data that has already been collected by others, although they may not necessarily have been analysed or published. Locating these sources and retrieving the information is a good starting point in any data collection effort.

Sources of existing available information include unpublished reports and publications in archives and libraries or in offices at various levels of public and private organizations, data bank of various government parastatals such as the Federal Bureau of statistics, Central Bank of Nigeria and International Development Agencies such as the World Bank, Food and Agriculture Organisation (FAO), etc.

In order to retrieve the data from available sources, the researcher will have to design an instrument such as a checklist or compilation sheet. In designing such instruments, it is important to inspect the layout of the source documents from which the data is to be extracted and then design the data compilation sheet in such a way that the items of data can be transferred in the order in which the items appear in the source document. This will save time and reduce error.

The advantage of using existing or secondary data is that collection is inexpensive. However, it is sometimes difficult to gain access to the records or reports required, and the data may not always be complete and precise enough, or too disorganised.

2. Observing

Observation is a technique that involves systematically selecting, watching and recording behaviour and characteristics of living beings, objects or phenomena. Observation can be participatory or non-participatory. In 'Participant Observation' the observer takes part in the situation he or she observes.

For example, a doctor hospitalised with a broken hip, who now observes hospital procedures 'from within' or a criminologist who pretends to be a prisoner and stays in the prison in order to observe and get first-hand information. In

'Non-participant Observation' the observer watches the situation, openly or concealed, but does not participate. For example wildlife studies in which the researcher stays inside the wild to observe the animals' activities.

Observations can give additional, more accurate information on the behaviour of people than interviews or questionnaires. They can also check on the information collected through interviews especially on sensitive issues such as alcohol or drug use, stigmatising diseases and crime.

3. Interviewing

An interview is a data collection technique that involves oral questioning of respondents, either individually or as a group.

Answers to the questions posed during an interview can be recorded by writing them down (either during the interview itself or immediately after the interview) or by tape-recording the responses, or by a combination of both. Interviews can be conducted with varying degrees of flexibility. The two extremes are high and low degree of flexibility.

High Degree of Flexibility Approach

This is useful if a researcher has as yet little understanding of the problem or situation he is investigating, or if the topic is sensitive. It is frequently applied in exploratory studies. The instrument used may be called an interview guide or interview schedule. The researcher will use a list of topics rather than fixed questions. For example when studying sensitive issues such as teenage pregnancy and abortions, the investigator may use a list of topics that may include how teenagers started sexual intercourse, the responsibility girls and their partners take to prevent pregnancy (if at all), and the actions they take in the event of unwanted pregnancies. The investigator should have an additional list of topics ready when the respondent falls silent, (e.g., when asked about abortion methods used, who made the decision and who paid). The sequence of topics should

be determined by the flow of discussion. It is often possible to come back to a topic discussed earlier in a later stage of the interview. The unstructured or loosely structured method of asking questions can be used for interviewing individuals as well as groups of key informants.

Low Degree of Flexibility Approach

Less flexible methods of interviewing are useful when the researcher is relatively knowledgeable about expected answers or when the number of respondents being interviewed is relatively large. Then questionnaires may be used with a fixed list of questions in a standard sequence, which have mainly fixed or pre-categorised answers.

Guidelines for Effective Use of Interview

1. The interviewer must create a congenial atmosphere of mutual respect, trust and understanding. He does this by:
   - giving the respondent the feeling that the experience he will gain from the interview will be pleasant and satisfying;
   - impressing on the respondent that the interviewer is reliable, responsible and trustworthy; and
   - convincing the respondent that the interview is important and his contributions are valuable and worthwhile.
2. A brief assertion of understanding and interest which is communicated to the respondent with comments such as "I see", "yes", "uh-huh", from the interviewer as he listens to the respondent.
3. An expectant pause. The simplest way to suggest to the respondent to say more may be pausing and wearing an expectant look or a nod of the head. This should not be used too frequently as it is likely to embarrass the respondent.
4. Repeat the question when the respondent appears not to have clearly understood it.

5. A neutral question or comment as a direct bid for more information. "Can you tell me more on that?" "Can you substantiate that statement?"

4. Administering written questionnaires

A written questionnaire is a data collection tool in which written questions are presented that are to be answered by the respondents in written form. A written questionnaire can be administered in different ways, such as by:

- Sending questionnaires by mail with clear instructions on how to answer the questions and asking for mailed responses;
- Gathering all or part of the respondents in one place at one time, giving oral or written instructions, and letting the respondents fill out the questionnaires; or
- Hand-delivering questionnaires to respondents and collecting them later.

The questions can be either open-ended or closed (with pre-categorised answers).

Guidelines for Effective use of Questionnaire

There is increasing evidence that a thoughtfully crafted introduction can be very important as it establishes a rapport. For example, it dispels any suspicion that the questioner works for the tax-gatherers; it introduces the themes and purpose of the survey. The introduction also develops the respondents' mindset, for example, by getting respondents to go over past events and recall situations that will inform the interview.

Transparency of intent should be established in the introduction and by following clear lines of questioning e.g. sections on household demographics, land tenure, crops and livestock. Within sections, it may be useful to follow a regular sequence of question types e.g. facts, practices, knowledge, attitudes and beliefs.

All questions to be included must be consistent with the objectives of the survey. It is often when the questionnaire is being planned that realisation dawns that the objectives have not been specified sufficiently precisely. To avoid rambling or obscure questions, put some issues and words in (i) a preamble, (ii) lists of permitted answers, or (iii) reiteration, confirmation, and extension of the first response.

Careful thought is needed to "profile" attitudes meaningfully. Often informants ought to participate directly in deciding the importance of profile elements. Open questions, which allow freer expression require disciplined data collection and may be difficult to summarise. Translators inexperienced in survey design may not appreciate the precision required in question wording, and with completion instructions and units of measurement. Therefore, look out for formally correct translations that are dialectally and culturally appropriate.

There is information from past studies to help with constructive approaches to many problems of questionnaire design. Ask those with relevant experience.

5. Focus Group Discussions (FGD)

A focus group discussion allows a group of 6 - 12 informants to freely discuss a certain subject with the guidance of a facilitator or reporter. The purpose of FGD is to obtain in-depth information on concepts, perceptions and ideas of a group. A FGD aims to be more than a question-answer interaction. The idea is that group members discuss the topic among themselves, with guidance from the facilitator.

6. Projective Techniques

When a researcher uses projective techniques, (s)he asks an informant to react to some kind of visual or verbal stimulus.

For example: An informant may be provided with a rough outline of the body and be asked to draw her or his perception of the conception or onset of an illness. Another example of a projective technique is the presentation of a hypothetical

question or an incomplete sentence or case-study to an informant ('story with a gap'). A researcher may ask the informant to complete in writing sentences such as:

- If I were to discover that my neighbour had TB, I would . . .
- If my wife were to propose that I use condoms, I would . . .

Such techniques can easily be combined with semi-structured interviews or written questionnaires. They are also very useful in FGDs to get people's opinion on sensitive issues.

7. Mapping and Scaling

- **Mapping** is a valuable technique for visually displaying relationships and resources. In a water supply project, for example, mapping is invaluable. It can be used to present the placement of wells, distance of the homes from the wells, other water systems, etc. It gives researchers a good overview of the physical situation and may help to highlight relationships hitherto unrecognised. Mapping a community is also very useful and often indispensable as a pre-stage to sampling.

- **Scaling** is a technique that allows researchers through their respondents to categorise certain variables that they would not be able to rank themselves. For example, they may ask their informant(s) to bring certain types of herbal medicine and ask them to arrange these into piles according to their usefulness. The informants would then be asked to explain the logic of their ranking.

Mapping and scaling may be used as participatory techniques in rapid appraisals or situation analyses.

Characteristics of Good Data Collection

- **Attributability**

Data can be traced to their source, e.g. by study number, sample

number and parameter. Unique identification of data pertaining to an individual study helps to prevent data mix-up.

- **Originality**

Raw data constitute the first recording of the observation. It should not be recorded on scraps of paper for transcription into a final form.

- **Promptness**

Data should be recorded immediately the operation is completed. It is not acceptable to make the recording sometime after the job has been finished, since memory may fail or become inaccurate, which may lead to data loss or faulty records.

- **Accuracy**

The raw data have to be a true representation of the observation and accuracy is thus absolutely central to the integrity of the study.

- **Legibility**

Data that cannot be read are useless and records that are difficult to decipher may raise doubts.

- **Indelibility**

Use indelible and water proof ink, to prevent deliberate tampering with data. Any changes to raw data should be made so as not to obscure the previous entry. Person responsible for the change should sign and date the change.

- **Storability**

The collected data represent the value (in time, resources and economic potential) of the research done. Therefore, the administration and physical placement of the storage facilities must be of good quality. Access to the storage facility should be limited to authorized personnel and facility must protect records from physical damage, interference and loss.

Data Collection Plan

A plan for data collection should be developed so that:

- one will have a clear overview of what tasks have to be carried out, who should perform them, and the duration of the tasks;
- one can organise both human and material resources for data collection in the most efficient way; and
- one can minimise errors and delays which may result from lack of planning (for example, the population not being available or data forms being misplaced).

It is likely that while developing a plan for data collection one will identify problems (such as limited manpower), which will require modification of the proposal. Such modifications might include adjustment of the sample size or extension of the period for data collection.

Stages in the Data Collection Process

Three main stages can be distinguished: (i) Permission to proceed; (ii) Data collection; (iii) Data handling.

Stage 1: Permission to Proceed

Consent must be obtained from the relevant authorities, individuals and the community in which the project is to be carried out. This may involve organising meetings at national or provincial level, at district and at village level. For clinical studies this may also involve obtaining written informed consent.

Most likely the principal investigator will be responsible for obtaining permission to proceed at the various levels. In many countries research proposals have to be screened for scientific and ethical integrity by national research councils.

Stage 2: Data Collection

When collecting our data, we have to consider:

- Logistics: who will collect what, when and with what resources.
- Quality control

1. Logistics of Data collection

WHO will collect WHAT data?

When allocating tasks for data collection, it is recommended that you list them first. Then you may identify who could best

implement each of the tasks. If it is clear beforehand that your research team will not be able to carry out the entire study by itself, you might plan to look for research assistants to assist in relatively simple but time-consuming tasks.

HOW LONG will it take to collect the data for each component of the study?

Step 1: Consider:

- The time required to reach the study area(s);
- The time required to locate the study units (persons, groups, records); If you have to search for specific informants, it might take more time to locate informants than to interview them; and
- The number of visits required per study unit. For some studies it may be necessary to visit informants a number of times, for example if the information needed is sensitive and can only be collected after informants are comfortable with the investigator or if observations have to be made more than once.

Step 2: Calculate the number of interviews that can be carried out per person per day (e.g., 4).

Step 3: Calculate the number of days needed to carry out the interviews. For example:

- you need to do 200 interviews,
- your research team of 5 people can do 5 × 4 = 20 interviews per day,
- you will need 200:20 = 10 days for the interviews.

Step 4: Calculate the time needed for the other parts of the study (for example, 10 days).

Step 5: Determine how much time you can devote to the study. Since the research team usually consists of very busy people, it is unlikely that team members can spend more than 30 working days on the entire study:

- 5 days for preparation (including pre-testing and finalising questionnaires),
- 20 days actual field work,
- 5 days data processing + preliminary analysis.

If the team has 20 days for fieldwork, as in the example above, it could do the study without extra assistance. However, if the research team has only five days available for the interviews, they would need five more research assistants to help complete this part of the study.

In WHAT SEQUENCE should data be collected?

In general, it is advisable to start with analysis of data that is already available. This is essential if the sample of respondents is to be selected from the records. Another rule of thumb is that qualitative research techniques (such as key informant interviews, focus group discussions) that are devised to focus the content of questionnaires for interviewing larger groups of informants should be carried out *before* finalisation of these questionnaires. FGDs designed to provide feedback on issues raised in larger surveys, should, logically, be conducted *after* preliminary analysis of the questionnaires.

To use time and transport efficiently, data to be drawn from different sources but in one locality should be collected at the same time. (For example, interviews with health staff in a Health Centre, observations of equipment available in the Health Centre and interviews with mothers living around the Health Centre should be scheduled together.)

WHEN should the data be collected?

The type of data to be collected and the demands of the project will determine the actual time needed for the data to be collected. Consideration should be given to:

- availability of research team members and research assistants;

- the appropriate season(s) to conduct the field work (if the problem is season-related or if data collection would be difficult during certain periods);
- accessibility and availability of the sampled population; and
- public holidays and vacation periods.

Bias in Information Collection

Bias in information collection is a distortion in the collected data so that it does not represent reality. Possible sources of bias during data collection include:

1. Defective instruments, such as:

Questionnaires with:

- fixed or closed questions on topics about which little is known (often asking the 'wrong things');
- open-ended questions without guidelines on how to ask (or to answer) them;
- vaguely phrased questions;
- 'leading questions' that cause the respondent to believe one answer would be preferred over another; or
- questions placed in an illogical order.
- Weighing scales or other measuring equipment that are not standardised.

These sources of bias can be prevented by carefully planning the data collection process and by pre-testing the data collection tools.

2. Observer bias

Observer bias can easily occur when conducting observations or utilising loosely structured group or individual interviews. There is a risk that the data collector will only see or hear things in which (s)he is interested or will miss information that is critical to the research.

Observation protocols and guidelines for conducting loosely structured interviews should be prepared, and training and practice should be provided to data collectors in using both these tools. Moreover it is highly recommended that data collectors work in pairs when using flexible research techniques and discuss and interpret the data immediately after collecting it. Another possibility -- commonly used by anthropologists -- is using a tape recorder and transcribing the tape word by word.

3. Effect of the interview on the informant

This is a possible factor in all interview situations. The informant may mistrust the intention of the interview and dodge certain questions or give misleading answers. For example: in a survey on alcoholism you ask school children: 'Does your father sometimes get drunk?' Many will probably deny that he does, even if it is true.

Such bias can be reduced by adequately introducing the purpose of the study to informants, by phrasing questions on sensitive issues in a positive way, by taking sufficient time for the interview, and by assuring informants that the data collected will be confidential.

It is also important to be careful in the selection of interviewers. In a study soliciting the reasons for the low utilisation of local health services, for example, one should not ask health workers from the health centres concerned to interview the population. Their use as interviewers would certainly influence the results of the study.

4. Information bias

Sometimes the information itself has weaknesses. Medical records may have many blanks or be unreadable. This tells something about the quality of the data and has to be recorded. For example, in a TB defaulter study the percentage of defaulters with an incomplete or missing address should be calculated.

Another common information bias is due to gaps in people's memory; this is called *memory* or *recall bias*. A mother may not remember all details of her child's last diarrhoea episode and of

the treatment she gave two or three months afterwards. For such common diseases it is advisable to limit the period of recall, asking, for example, 'Has your child had diarrhoea over the past two weeks?'

All these potential biases will threaten the validity and reliability of your study. By being aware of them it is possible, to a certain extent, to prevent or reduce them. If the researcher does not fully succeed, it is important to report honestly in what ways the data may be biased.

Ensuring Quality in Data Collection

It is extremely important that the data we collect are of good quality, that is, reliable and valid. Otherwise we will come up with false or misleading conclusions. There are a number of possible measures that can be taken to prevent and to partly correct these distortions.

Measures to help ensure good quality of data:

- **Prepare a field work manual for the research team as a whole, including:**
  - Guidelines on sampling procedures and what to do if respondents are not available or refuse to co-operate.
  - A clear explanation of the purpose and procedures of the study which should be used to introduce each interview, and
  - Instruction sheets on how to ask certain questions and how to record the answers.
- **Select your research assistants, if required, with care.** Choose assistants that are:
  - from the same educational level;
  - knowledgeable concerning the topic and local conditions;
  - not the object of study themselves; and
  - not biased concerning the topic (for example, health staff are usually not the best possible interviewers for a study on alternative health practices).
- **Train research assistants carefully in all topics covered** in the field work manual as well as in interview techniques and make sure that all members of the research team master interview techniques such as:
  - asking questions in a neutral manner;
  - not showing by words or expression what answers one expects;
  - not showing agreement, disagreement or surprise; and
  - recording the answers precisely as they are provided, without sifting or interpreting them.
- **Pre-test research instruments and research procedures** with the whole research team, including research assistants.
- **Take care that research assistants are not placed under too much stress** (requiring too many interviews a day; paying per interview instead of per day).
- **Arrange for ongoing supervision** of research assistants. If, in case of a larger survey, special supervisors have to be appointed, guidelines should be developed for supervisory tasks.
- **Devise methods to assure the quality** of data collected by all members of the research team. For example, quality can be assured by:
  - requiring interviewers to check whether the questionnaire is filled in completely before finishing each interview;
  - asking the supervisor to check at the end of each day during the data collection period whether the

questionnaires are filled in completely and whether the recorded information makes sense; and

- having the researchers review the data during the data analysis stage to check whether data are complete and consistent.

Use of Standard Operating Procedures (SOPS) in experimental data collection

These are documents which describe activities of a repetitive, routine nature in a very detailed manner. SOPs should be immediately available to all individuals performing the respective tasks, and must be followed rigorously.

SOP ensures, among other things:

- Standardization, consistent procedures (person-to-person, test-to-test variability minimized)
- Continuity in the event of personnel turnover.

Use of Pilot Study

This is a major approach to ensure quality in research. Pre-test or pilot study usually refers to a small-scale trial of particular research components. It is the process of carrying out a preliminary study, going through the entire research procedure with a small sample.

WHY do we carry out a pre-test or pilot study?

- A pre-test or pilot study serves as a trial run that allows us to identify potential problems in the proposed study. Although this means extra effort at the beginning of a research project, the pre-test and/or pilot study enables us, if necessary, to revise the methods and logistics of data collection before starting the actual fieldwork. As a result, *a good deal of time, effort and money can be saved* in the long run. Pre-testing is simpler and less time-consuming and costly than conducting an entire pilot study. Therefore we will concentrate on pre-testing as an essential step in the development of research projects.

WHAT aspects of your research methodology can be evaluated during pre-testing?

1. Reactions of the respondents to the research procedures can be observed in the pre-test to determine:
   - availability of the study population and how respondents' daily work schedules can best be respected;
   - acceptability of the methods used to establish contact with the study population;
   - acceptability of the questions asked; and
   - willingness of the respondents to answer the questions and collaborate with the study.
2. The data-collection tools can be pre-tested to determine:
   - Whether the tools you use allow you to collect the information you need and whether those tools are reliable. You may find that some of the data collected is not relevant to the problem or is not in a form suitable for analysis. This is the time to decide not to collect this data or to consider using alternative techniques that will produce data in a more usable form.
   - How much time is needed to administer the interview guide/questionnaire, to conduct observations or group interviews, and/or to make measurements.
   - If there is any need to revise the format or presentation of interview guides/ questionnaires, including whether:
     - The sequence of questions is logical;
     - The wording of the questions is clear;
     - Translations are accurate;
     - Space for answers is sufficient;
     - There is a need to pre-categorise some answers or to change closed questions into open-ended questions;
     - There is a need to adjust the coding system; and
     - There is a need for additional instructions for interviewers (e.g., guidelines for 'probing' certain open questions).

3. Sampling procedures can be checked to determine:
   - Whether the instructions concerning how to select the sample are followed in the same way by all staff involved.
   - How much time is needed to locate individuals to be included in the study.
4. Staffing and activities of the research team can be checked, while all are participating in the pre-test, to determine:
   - How successful the training of the research team has been;
   - What the work output of each member of the staff is;
   - How well the research team work together;
   - Whether logistical support is adequate; and
   - The reliability of the results when instruments or tests are administered by different members of the research team.
   - Whether staff supervision is adequate.

The pre-test can be seen as a period of extra training for the research team in which sensitivity to the needs and wishes of the study population can be developed.

5. Procedures for data processing and analysis can be evaluated during the pre-test. Items that can be assessed include:
   - Appropriateness of data master sheets and dummy tables and the ease of use;
   - Effectiveness of the system for quality control of data collection;
   - Appropriateness of statistical procedures (if used); and
   - Clarity and ease with which the collected data can be interpreted.
6. The proposed work plan and budget for research activities can be assessed during the pre-test. Issues that can be evaluated include:
   - Appropriateness of the amount of time allowed for the different activities of planning, implementation, supervision, co-ordination and administration.
   - Accuracy of the scheduling of the various activities.

When do we carry out a pre-test?

You might consider:

- Pre-testing at least your data collection tools, either during the workshop, or, if that is impossible, immediately thereafter, in the actual field situation.
- Pre-testing the data collection and data analysis process 1-2 weeks before starting the fieldwork, with the whole research team (including research assistants, if required) so that you have time to make revisions.

Which components should be assessed during the pre-test?

1. **Pre-test during the workshop**

Depending on how closely the pre-test situation resembles the area in which the actual field work will be carried out, it may be possible to pre-test:

- The reactions of respondents to the research procedures and to questions related to sensitive issues.
- The appropriateness of study type(s) and research tools selected for the purpose of the study (e.g., validity: Do they collect the information you need? and reliability: Do they collect the data in a precise way?).
- The appropriateness of format and wording of questionnaires and interview schedules and the accuracy of the translations.
- The time needed to carry out interviews, observations or measurements.
- The feasibility of the designed sampling procedures.
- The feasibility of the designed procedures for data processing and analysis.

Even if you cannot assess all these components fully, the field experience will provide information that will be quite valuable to you in reviewing the methodological aspects of your proposal and in planning your work plan and budget.

2. **Pre-test in the actual research area**

All the issues mentioned above will have to be thoroughly reviewed during a pre-test in the actual field situation. Other

issues, such as the functioning of the research team, including newly recruited and trained research assistants, and the feasibility of the work plan, can only be tested in the research area. An important output of the pre-test should be a fully developed work plan. If choices have to be made as to what to include in the pre-test, the following considerations may be helpful:

- **What difficulties do you expect in the implementation of your proposal?**

Think of possible sources of bias in data collection techniques and sampling and ethical issues you considered during the preparation of your plan for data collection. Can some of these potential problems be overcome by adapting the research design?

- Inexperience with a certain data-collection technique is also a reason to include it in the pre-test.
- Which parts of your study will be most costly and time-consuming? Questionnaires used in large surveys, for example, should always be tested. If many changes are made, the instruments should be pre-tested again. If an interview guide or questionnaire has been translated into a local language, the translated version should be pre-tested.

It is highly recommended that you analyse the data collected during the pre-test right away. Then finalise and adjust the master sheets, if necessary. Make totals for each variable included in the master sheets. Fill in some dummy tables and prepare all the dummy tables you need, considering your research objectives.

Do all this even if you plan to analyse the data by computer. You will detect shortcomings in your research tools that you can still correct!

Who should be involved in the pre-test or pilot study?

- The research team, headed by the principal investigator.
- Any additional research assistants or data collectors that have been recruited.

How long should the pre-test or pilot study last?

The time required for a pre-test or pilot study will be determined by a number of factors:

- The size and duration of the research project. (The longer the study will take, the more time you might reserve for the test run.)
- The complexity of the methodology used in the research project.

Keep in mind that this is the last chance you will have to make adjustments which will help to ensure the quality of your fieldwork. If you have a 20 day field work period you might reserve at least 3-5 days for pre-testing your data collection tools, analysing the results of the pre-test, finalising your tools and elaborating the work plan.

Stage 3: Data Handling

Once the data have been collected and checked for completeness and accuracy, a clear procedure should be developed for handling and storing them. One of the major methods of data handling is archiving which involves data collectors keeping perhaps on CD, all the non-ephemeral material relating to their efforts to acquire information. Obvious components of such a record include:-

- data collection instruments,
- raw data,
- metadata recording the what, where, when, and other identifiers of all variables,
- variable names and their interpretations, and labels corresponding to values of categorical variables,
- query programs used to extract analysis files from the database,
- log files defining the analyses, and reports.

Often geo referencing information, digital photographs of sites and scans of documentary material are also useful. Participatory

village maps, for example, can be kept for reference as digital photographs. Surveys are often complicated endeavours where analysis covers only a fraction of what could be done. Reasons for developing a good management system, of which the archive is part, include:-

- keeping the research process organised as it progresses;
- satisfying the sponsor's (e.g. DFID's) contractual requirement that data should be available if required by the funder or by legitimate successor researchers;
- permitting a detailed re-analysis to authenticate the findings if they are questioned;
- allowing a different breakdown of results e.g. when administrative boundaries are redefined;
- linking several studies together, for instance in longer-term analyses carrying baseline data through to impact assessment.

Other data handling procedure includes:

- Numbering of the questionnaires and other research tools. Decision has to be made if this should be done at the time of the interview or at the time the questionnaires are stored.
- Identify the person responsible for storing data and the place where it will be stored.
- Decide how data should be stored. Record forms should be kept in the sequence in which they have been numbered.

References

Dawson B. and Trapp R.G (2001): *Basic and Clinical Biostatistics*. Lange Medical Books/McGraw-Hill, Medical Publishing Division, New York.

IDRC (2008): Designing and conducting health systems research projects- modules 10, 10A, 10C, 12 and 14. /www.idrc.ca/en/ev-56622-201-1-DO_TOPIC.html

Lumley J.S.P and W. Benjamin (1994): *Research: Some Ground Rules*. Oxford New York Tokyo, Oxford University Press.

National Academy of Sciences (1995): *On Being A Scientist: Responsible Conduct in Research*. National Academy Press.

TDR (2001): *Quality Standards in Basic Biomedical Research*. 16pp.

Statistical Services Centre, The University of Reading, UK. (2000): Concepts Underlying the Design of Experiments.

Statistical Services Centre, The University of Reading, UK (2000): *Guidelines for Planning Effective Surveys*.