

Household Data Collection for Socio-Economic Research in Agriculture: Approaches and Challenges in Developing Countries

Dickson M. Nyariki

*Department of Land Resource Management and Agricultural Technology,
University of Nairobi, P.O. Box 29053-00625, Nairobi, Kenya
E-mail: dicksonnyariki@yahoo.com*

KEYWORDS Socio-Economic Studies. Agriculture. Developing Countries. Rural Studies

ABSTRACT In this article, household data collection procedures for socio-economic research in agriculture under conditions prevalent in developing countries are described. Using relevant examples for Kenya, the sources and types of data, the procedures adopted in surveys and other forms of data collection, including focus group discussions and questionnaire use, are discussed. Also elaborated are the processes of choosing and training enumerators, pilot survey procedures, and field interviews. The challenges faced by researchers in data collection for agriculture-related studies in a developing country context are pointed out.

INTRODUCTION

Household data are frequently used in the analysis of various socio-economic phenomena relevant to agriculture (including livestock production). However, the results of household or rural studies are as good as the process involved in collecting the required information and data. If inappropriate methods to collect information are used, the analysis, results, conclusions and the accompanying recommendations would be erroneous and, therefore, of little or no value to the end-user—the policy maker and implementer.

The kinds of data collected and their sources need to be in accordance with the objectives and hypotheses of the study. The first step in data collection, therefore, is to carefully consider the aspects of clarity, specificity, technical feasibility, and cost-effectiveness of the approaches adopted to address the issue or topic of interest to the researcher. The study issue will have a conceptual and theoretical basis, backed by research objectives, hypotheses, and relevant analytical procedures. Thus, the data needs will be determined by all the pertinent aspects of the research study.

The data collection process would, as its main components, consist of sampling, a decision on the sampling design, determination of the sample size, selection of collection tools, and field surveys, among others. The purpose of this article is to discuss and elaborate on these components, and the methods and approaches used in

household data collection, particularly for agricultural and household economics research. Challenges and limitations faced by researchers under conditions of a developing country are highlighted and ways to circumvent them are suggested throughout the article. It is expected that this article would be of interest to research students, academicians and other researchers in the private and public arenas.

SOURCES AND TYPES OF SOCIO-ECONOMIC DATA

Two types of data are recognised. These are *primary* and *secondary* data. Primary data are collected through questionnaire interviews, focus group discussions (FGDs), and other participatory approaches. Primary data are raw data while secondary data are already processed and may be in the form of reports, publications, and or stored in any other forms—soft or hard. The focus in this article is primary data collection.

Data may also be categorised as *qualitative* and *quantitative* as well as *cross-sectional* and *time series*.¹ They may be based on years, seasons, months or shorter periods. The decision on the frequency of collecting each datum would usually be influenced by the nature of the datum and the purpose of collection.

By recognising that people's recall deteriorates through time, it is recommended that household data spanning back more than one year should not be solicited, unless they are

recorded. In some cases, to cross-check quantitative data, respondents are asked to report information in a rank order form. Some respondents may be unable to give quantitative information, simply because it does not exist.

SAMPLING

To obtain data for socio-economic research a *sample* is usually required. The reason is that the population may be so big as to make it difficult for one to get to every individual. One must therefore sample the population.

A sample is a *subset of a population* from which to obtain information. The sample can then be used to *generalize* about the population. However, generalization is only possible and appropriate if the sample is *representative* of the target population. The population may be *homogeneous* or *heterogeneous*. Homogeneous populations do not present problems in sampling; a small random sample would suffice as opposed to heterogeneous populations. A homogeneous population is like a well mixed mug of porridge with sugar or a meal of boiled maize and beans, the latter known in some local dialects in Kenya, where it is popularly consumed, as *Githeri* or *Nyoyo*. If you scoop any spoonful of such porridge or any bowl of such *Githeri*, you are bound to obtain the same consistency or composition; on the one hand in terms of the amount of sugar and flour or, on the other hand, in terms of maize and beans. If a population is unevenly distributed, some form of *mixing* is required prior to initiating sampling. This is the *sampling design*. Thus, the distribution of a population will affect the decision on the kind of sampling design to be adopted.

Sampling Difficulties in Developing Countries

One of the main reasons research data are likely to be inaccurate in developing countries is because of difficulties in carrying out proper sampling. The basic rule in research is that, for a sample to truly reflect the characteristics of the target population, random allocation has to happen such that every member of the population has an equal chance of appearing in the sample. In practice, this is difficult to achieve and researchers end up compromising this tenet for practical purposes. The bigger the compromise, the larger is the sampling error.

In developed countries, such as Europe and USA, household data collection for research relies on an advance in development that does not obtain in developing countries such as Kenya—the postal code. For example, unlike most countries in the developing world, every household in Britain has a postal code so that when researchers conduct household surveys, they can obtain an accurate list of all households. Furthermore, mobile and fixed telephony is advanced, readily available and accessible to all. Since almost everybody has a telephone line or its mobile variety at home and at work, this enables a researcher to access accurate data on the population to be surveyed. After this, appropriate random samples are selected using computer packages to allocate random numbers.

In contrast, in a country like Kenya very few households in the rural areas can be reached by telephone or cell-phone. In addition, the terrain is ragged and the road network is poor, both in terms of condition and connectivity. It is therefore very difficult to randomly sample the population. Therefore, researchers would sample the administrative areas of study but when it comes to the actual selection of the study area and the sample to be studied, logistical and geographical difficulties set in, and the selected sample is far from being random. Many researchers would resort to physical identification methods, which are usually inaccurate. If not, they would find short-cuts to this.

Selection of Sampling Design

A *sampling design* is the procedure followed in sampling. There are several such designs, the choice of which should be dictated by the factors already pointed out. The common designs and procedures in socio-economic research in a developing country are briefly discussed.

Simple Random Sampling. This is a procedure where each of the individuals has an equal chance of being picked. To carry out simple random sampling a *sampling frame* is required. This is an ordered list of individuals in the *population of interest* to the researcher. In developing a sampling frame, the researcher may carry out a complete census, which could be very expensive to do, or may choose other means such as using a telephone book or telephone numbers. The latter is not common in agricultural socio-economic research because of inadequate telephone

connectivity among households in the rural areas of developing countries.

In the case of socio-economic research in an African country such as Kenya, the assistance of the village headmen or chiefs, or some other local leader, is normally sought. A list of all the households is obtained from which a sample frame is drawn. This may depend on administrative boundaries. In some cases, using these boundaries is not desirable as we may be dealing with agro-ecological zones that traverse the boundaries. Thus, before sampling, the researcher must familiarise with the area and population of interest. To help in this, *reconnaissance surveys* and literature reviews are carried out.

Stratified Random Sampling. Because of common heterogeneity among agricultural household populations, it is usually desirable to undertake population stratification. The population is first classified into relatively homogeneous groups. After this a *proportional sample from each stratum* is taken at random depending on the number in each stratum. If randomisation is not done, this form of population stratification is referred to as *quota sampling*.

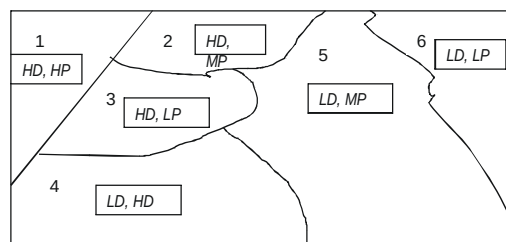
Cluster Sampling. Cluster sampling, which should not be confused with stratified random sampling, entails dividing the population into several groups that are regarded as belonging to different subgroups, e.g., population distribution, ecological potential or farming systems (see Figure 1), and then taking random samples separately from these groups. It is usually economical to sample specific groups (clusters) and then obtain data from each eligible unit sampled (household). Note that although the population is divided into groups for both cluster sampling and stratified random sampling, the two techniques differ in that in stratified random sampling, a simple random sample within each group is taken, while in cluster sampling a simple random sample of groups is taken and then all units (households) are sampled within the selected groups (Hoinville and Jowell 1978; Scheaffer et al. 1979).

Further cost reduction and convenience may be achieved if a list of units in the population of interest can be available. This may be done by selecting a 'name' of the units near the beginning of the list and then selecting, say, the tenth or twentieth, or any other name thereafter, depending on the number of names in the list and the size of the sample required. This is called *systematic*

sampling. Since research is not conducted with unlimited resources, cluster sampling is more often than not necessary, especially in a developing country (like Kenya) because of poor infrastructure and difficult terrain (Casley and Lury 1987).

Figure 1 shows that a study area has been divided into clusters based on geographical areas, by considering various demographic and ecological characteristics. These clusters are then numbered. After this, they would be randomised and the chosen ones picked and studied. The researcher may decide to sample these numbers again before studying. This is called *two-stage* or, more generally, *multi-stage random sampling*.² Cluster sampling may also be based on administrative areas, e.g., sub-location, location, division, district, or any other administrative region.

If, for example, after the selection of clusters, the researcher obtains a list of 1,500 farm-household names in total, the names would then be numbered to form a sampling frame. Assuming the sample size determined was 100 units, every 15th farmer would be selected to obtain the 100 farm-households—by applying systematic sampling to the frame. The starting number in this case would lie between 1 and 15, which, as may be noted, depends on the number of farm-families and the sample size required. This number should also be randomly selected. This kind of sampling reduces expenses and is therefore appropriate where the availability of funds is highly constrained, as would be the situation in developing countries. For more detailed discussions on the designs of sample surveys, see, for example, Freund and Williams (1983), Fowler (1984), Devereux and Hoddinott (1992), and Casley and Lury (1987).



Key: HD=High (population) density, LD=Low density
HP=High (land) potential, MP=Medium potential
LP=Low potential

Fig. 1. An illustration of cluster sampling

Determination of Sample Size

The number of sample units to be selected from a study population is usually a moot question among researchers. When the population size is known or roughly so and the researcher has taken care of the heterogeneity problem, any number (equal to or) greater than the statistically *large sample* (of 30 sample units) may be appropriate (Freund and Williams 1983). However, in addition to taking into consideration the statistical requirement to have a minimum size of 30 sample units, the ultimate size arrived at must bring into view the possibility of non-response (which, by 'playing it safe', may be given a 40–50% chance), limited financial outlays expected in a developing economy, the nature of the research, and time—as most research categories are time-bound. Furthermore, the relative difficulties of accessing the study area due to the nature of the terrain and the level of development of the infrastructure must be assessed and taken into account. Taking all these factors together, larger samples are usually desirable, but care must be taken to ensure that the samples are not so large as to reduce the resources available per unit and to compromise the quality of the data collected.

In more stringent statistical considerations and data collection procedures, sample sizes are systematically determined, by employing the *Probability Proportional to Size* (PPS) sampling method. In agricultural socio-economic research, this kind of sampling, usually with a 95% confidence level and an error margin of less than 10%, is considered representative, and has statistical significance that facilitates the collection of information on socio-economic indicators under investigation and the making of valid statistical inferences. Where simple random sampling is involved, the sample size is calculated using a standard formula, thus (Freund and Williams 1983):

$$N = z^2 (pq) / d^2 \quad (1)$$

Where: N = sample size
 z = statistical certainty usually chosen at 95% confidence level, that is, $z = 1.96$
 for an error risk of 5%
 p = estimated level/coverage to be investigated, usually $p = 0.5$ is chosen
 $q = 1 - p$
 d = precision desired, expressed as a fraction of 1, usually $d = 0.1$ is chosen

For the confidence interval and the rest of the values given above, the sample size required for the study is calculated (by substituting the values in equation 1) as:

$$\begin{aligned} N &= (1.96)^2 (0.5 \times 0.5) / (0.1)^2 \\ &= (3.84) (0.25) / 0.01 \\ &= 96 \end{aligned}$$

The *Sampling Interval* (SI) is calculated by dividing the total population by the sample size ($N = 96$). If, for example, the study area has n distinct sub-areas (say districts) with different population sizes and one wishes to survey all of them, the sum total of the population will be calculated thus:

$$\sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n \quad (2)$$

Where: X_i = Population in sub-area i
 n = number of sub-areas (districts)

The total population derived as shown in equation 2 is divided by the sample size to determine the SI, as follows:

$$SI = \frac{\sum_{i=1}^n X_i}{96}$$

DATA COLLECTION TOOLS AND PROCEDURES

Data collection tools are the *instruments* used for obtaining data from a population or sample. The procedures, on the other hand, are the processes or mechanisms adopted, using the tools, to solicit the data. The basic tools and procedures in socio-economic research in agriculture are questionnaires and *Focus Group Discussions* (FGDs). The preparation and administration of appropriate questionnaires and conduction of effective FGDs are therefore important.

Questionnaire Preparation

After samples have been selected, interviews may be carried out using a questionnaire. A questionnaire is a set of questions organised in a systematic way for the purpose of eliciting information from respondents. It may be (fully) structured or semi-structured, or may simply be a checklist of questions, to provide the required data.

A draft questionnaire taking into account the objectives and the hypotheses of the research is constructed before setting out to the field. The questionnaire frequently takes the three common forms: it would contain dichotomous, multi-choice and open-end(ed) questions. This is

usually necessary because of the diverse issues that are often investigated.

A dichotomous questionnaire is the simplest as it requires the respondent to choose between one of two responses. Although dichotomous questions are easy to construct and the corresponding answers quick to record, their main disadvantage is that they usually oversimplify issues and do not provide room for compromise. Therefore, more often than not, a multiple-choice type questionnaire is desirable. As for the open-ended questionnaire, a person is allowed the most freedom of response. However, as would be expected, the disadvantage of open-ended questionnaires is the difficulty of classifying the results obtained.

In preparing a socio-economic research questionnaire, an effort should be made to ensure that each question is simple and is phrased in a manner that would imply the same meaning to all that are to be interviewed. In other words, questions that would carry more than one meaning must be avoided. Leading questions should also be avoided as they usually suggest the answer the interviewer wants to hear, and the respondent may agree with the interviewer simply because that is the easiest or expected response.

Sequencing of questions is important. The more sensitive questions, such as those inquiring about family size, age, property ownership, and scholarship, should appear later in the questionnaire. They are better held back until such time when the interviewer has struck a rapport with the interviewee. The special nature of questions requiring numerical responses must also be recognised. Questions should be constructed in a way that allows adequate room to make considered choices, so as to avoid forcing answers. The possibility for no response must be borne in mind. For example, some people may not remember their exact age, in which case the appropriate question would be that which gives a range of ages.

One of the main features of a good questionnaire is short and few questions, including only questions pertinent to the objectives of the study. This is to avoid people becoming bored after answering an unending list of questions, which may also lead to incorrect answers (Hoinville and Jowell 1978). As a rule of thumb, interviews should take between 45 minutes and one hour. If for some reason an interview must take more than one hour before the questionnaire can be fully filled, it should be put off and a fresh appointment made with the interviewee for another day to complete the

exercise. It is only if the interviewee showed enthusiasm to continue, and the researcher or enumerator was satisfied that this would not impair the accuracy of the answers given, might the interview be allowed to proceed.

Focus Group Discussions

Focus Group Discussions are a commonly used qualitative approach to data collection. The importance of qualitative approaches in understanding social realities has been recognised by many social scientists. This has arisen from the need to question the adequacy of an exclusively quantitative approach in explaining changes in social phenomena. The FGD method is in fact one of the most popular among socio-economic research methods and is extensively used in social and behavioural studies. Thus, in addition to the various qualitative techniques, including in-depth interviews, observations, case studies and narratives, the FGD method is an important technique (Odimegwu 2000).

A *focus group session* is a discussion in which a number of respondents (best if about 10 of them) under the guidance of a moderator, talk about topics that are believed to be of special importance to the investigator (Mulwa and Nguluu 2003). In recent years, there has been a notable growth in the use of focus group session research to gain insight into the dynamic relationships of attitudes, opinions, motivations, concerns and problems related to current and projected human activity (Odimegwu 2000). FGDs are widely used in household and agricultural economics, private industry or business where they are considered essential to understand the psychological and behavioural underpinnings of producer and consumer behaviour and to identify ways and means to influence this behaviour.

Usually more than one group session is conducted to ensure adequate coverage. The discussions capitalize on group dynamics and allow a small group of respondents to be guided by a skilled moderator into increasing levels of focus and depth on the key issues of a research topic. The discussions are conducted as open conversation in which each participant may comment, ask questions, or respond to comments of others (Mulwa and Nguluu 2003).

The FGD is basically a method in which a moderator, with the help of predetermined guidelines, stimulates free discussions among

participants on the subject of the inquiry. Participants are chosen from some specific target groups whose opinions and ideas are germane to the investigators. The participants are purposively selected and it is recommended that they be of the same sex, age-set, socio-economic background, profession or interest.

The order in which topics are covered is flexible but generally the discussions start with more general issues and slowly flow into more specific ones. Towards the end, a few probing questions are asked to reveal more in-depth information or to clarify earlier responses. If the topic is of a sensitive nature, the moderator should be of the same sex with the participants as well. It is also desirable that the participants do not know the details of the subject of discussion in advance. The session should be held in a natural setting and be conducted in a relaxed manner. There should also be a note taker who should not participate in the discussion. The note taker should know the subject and the objectives of the research and is expected to be well trained in observing and noting nonverbal group feedback, for example, facial expression and side talks. He should be able to record and translate the complete discussion based on his notes.

The homogeneity of the group setting and the open-ended nature of the questions encourage the participants to feel free from various constraints to which they are subjected during individual interviews. It is believed that they express their views openly and spontaneously. Well-guided and unbiased moderation helps the participants to build confidence and interact. This interaction stimulates memories and feelings and in turn leads to full and comprehensive discussions of the topic at hand.

There are several limitations associated with FGDs, however. One, a group setting is not always ideal for encouraging free expression. Sometimes the group can be intimidating and as a result inhibit discussion. Two, care must be taken in preparing transcripts from notes. Chances of introducing errors are particularly high if the discussion has to be translated from the local language to the language of the investigator. Three, the samples are usually small and purposively selected, and therefore do not allow much generalisation to larger populations. The chances of introducing bias and subjectivity into the data are also high.

While the FGD method can provide plausible

insights and explanations, one should not extrapolate from focus group discussions to the distribution of responses in a population. There is concern that focus group research outcomes are being misused as quick and easy claims to the validity of the research and are not subjected to further quantitative tests.

FIELD SURVEYS

Field surveys may include preliminary survey procedures that include *Rapid Rural Appraisal* (RRA), other participatory approaches, and pilot surveys. These are conducted prior to the commencement of data collection for the main sample.

Rapid Rural Appraisal

Normally, during the questionnaire pre-testing exercise, RRA is carried out in the study area. In this particular circumstance, this procedure, which is known in some disciplines as *Rapid Assessment Procedure* (RAP) or *Informal Agricultural Survey* (IAS), would essentially be a reconnaissance survey. In the case of socio-economics in agriculture, this would be done to establish, among others, farmers' operations concerning crops grown, livestock kept, foodstuffs consumed and resources used. The appraisal may involve walking or driving through the survey area to observe the fields and the animals, meeting elders and talking to farmers and questioning them informally about the mentioned operations.

There has been hot debate on the merits and demerits of RRA and related 'methodologies' like *Participatory Rural Appraisal* (PRA), *Participatory Learning and Action Research* (PLAR), and *Participatory Action Research* (PAR). There has also been no clear consensus on what they entail, whether they are approaches, methodologies or data collection tools (see, for example, Chambers, 1983; Chambers, 1997; Neefjes, 2000). What may be clear, however, is that the PRA, PLAR and PAR are approaches that entail greater and longer involvement and participation of and learning by the local communities than the RRA, because they go beyond simple rural appraisals.

Pilot Survey

It is recommended that a research question-

naire is pre-tested in a pilot survey involving a few units (households), say 10, from the target population before it is used in the main survey. The 10 households selected for questionnaire pre-testing should belong to the same area of survey but must not come from the main sample. The core reasons for pre-testing the pre-field questionnaire are to decide whether or not to exclude or modify (rephrase) some of the questions. This is done to ensure that the final questionnaire has only relevant and appropriately phrased questions to be put to the interviewees from the main sample.

Personal Interviews

To avoid exhausting the enumerators or the researcher, leading to a loss of concentration that could reduce the quality of data collected, a maximum of five interviews may be conducted—three in the morning and two in the afternoon to evening. Probably, on average, three respondents per day should suffice to ensure the interviewers are kept keen during the interview and to keep the cost of the exercise reasonably low. During these interviews, initial visits would normally be made with the intention of making appointments that are appropriate to the interviewee.

In agriculture, field visits at different times or seasons of the year are helpful in tracking agricultural production and consumption patterns, for example, and to reduce the recollection period for purposes of accuracy. They also provide time-series data based on seasons, and avoid snapshot data and results. If, for instance, the issue is about foodstuffs prepared and consumed by a household, questions would be put to the farmer concerning whether the 24-hour period or week preceding the field visit was 'normal' or not (Bouis *et al.* 1992; Webb and von Braun, 1994). This is likely to result in more accurate information than, say, a week or longer period of recollection.

Repeat Surveys

Repeat surveys are defined as the collection of current longitudinal data through multiple interviews over a defined time period (Olsen, 1992). The main reasons for repeat surveys in agricultural research are, one, to collect detailed information on farm-household activities over a relatively long period (over one year), because the period is too long for accurate recall without

repeat surveys, and, two, to obtain data on changes in variables that are expected to exhibit variability over several seasons. As indicated by Olsen, this latter reason is particularly relevant in agriculture and other similar studies where research is conducted in an area characterised by seasonality and or drought. If these factors are ignored they are likely to make the short-term fieldwork unrepresentative.

By spending a year or longer in the field, and by using repeat surveys, it is possible to capture seasonal differences in, for example, household production activities, consumption and nutrition, and patterns of buying and selling. In other words, if repeat surveys are not done in an area that experiences seasonality, a distorted snap-shot would be the result. This is especially likely in cases where the research environment is characterised by both seasonality and droughts, frequently associated with atypical agricultural seasons or years.

Repeat surveys have a number of limitations. Chief among these is that there are bound to be changes in the numbers and characteristics of households due to mobility and morbidity or deaths, and or resource transformation. The longer the period between surveys, the more these changes become apparent. Some of the households may lose their heads in death and some may have their households subdivided. This latter situation occurs where mature sons 'move out' to start their own households. This means that if the new households were to be interviewed, the number of households would increase. If the younger households are ignored, it means that some of the resources, such as land and even labour are reflected as lost. Land may also be 'lost' through sales. Either way, there would be an analytical problem to be tackled.

Other than the difficulties related to the analysis and interpretation of data, repeat surveys involving long periods of study, spanning several years, may pose additional problems. One such problem is the uncertainty of future prices and costs. A researcher is thus likely to experience unanticipated escalation of survey costs, incurred mainly because of the high inflation rates pervasive in developing countries.

Resources would be in some cases gained when households purchase extra land and members become more mature and or acquire more skills, so that the quality and quantity of labour goes up. Emigration of entire households may

also take place. The movements and sharing up of land would lead to a reduction in the total size of land occupied by the 'original' households, if the new households are not interviewed. Whichever way such differentiation is treated, we may not be dealing with exactly the same households in the two periods. Such are some of the limitations of repeat surveys.

For purposes of illustrating repeat surveys in agriculture and related studies, let us take a specific example of a PhD research study carried out in Kenya (Nyiriki 1997). The objective of this study was to test hypotheses concerning household food availability and access, incomes and poverty levels, and resource availability and use in the semi-arid areas of Kenya. As implied, the study entailed the analysis of household data (Table 1). These data were to be obtained mainly through surveys conducted in one of the semi-arid districts as a 'representative' or 'case' of semi-arid areas in Kenya. The data were to be gathered through repeat surveys. In this district, households are involved in small-scale farming.

Table 1 provides an example of the methods and frequency of data collection in repeat interviews for the study. As may be noted, a decision had to be made on which variable whose data collection required simple observations by the researcher, recall by the interviewee, extraction from records, or a combination of two or more methods. Also, certain data were collected once per visit and others only once during the full period of the study. All these would be dependent upon how the research study was formulated.

Identification and Training of Survey Enumerators

The procedure of obtaining data by personal interviews often requires the interviewer to ask prepared questions in a formal questionnaire, discussed above, and to record the respondent's answers. The main advantage of this kind of interview is that people will usually respond when confronted in person. Furthermore, the interviewer will be able to note specific reactions and therefore eliminate any misunderstanding that may arise from the questions being asked.

The major limitations of personal interviews, apart from the high costs that are commonly involved, are related to the interviewers themselves. If they are not thoroughly trained and closely inspected, they may deviate from the

required protocol, thereby introducing bias in the sample data collected. Also, any movement, statement, or facial expression by the interviewer could affect the response obtained. For example, a leading question such as, 'Don't you think dipping your cattle once a month is inappropriate?' might tend to elicit an answer more likely positive than negative. In addition, errors introduced in recording could lead to wrong results. These are some of the reasons why the cost spent on well trained survey enumerators is justified.

Enumerators should be adequately educated, preferably with a minimum of an 'O' level certificate. The enumerators should, of necessity, be recruited from the local community in which the study is being conducted, to circumvent language barrier, so that the information obtained would be as accurate as possible. The local enumerators, who are members of the villages surveyed, would also be useful in identifying the sample households selected.

The unreliability of enumerators, who may occasionally drop out half-way through the interviews, is a source of increased survey costs. This is often caused by poverty and the concomitant searches for better paying work elsewhere. If this occurs, it would necessitate the recruitment and training of new enumerators in the middle of the study. This would, therefore, make it quite difficult to keep within the survey schedule and the allocated budgets.

Table 1: An example of data collection through repeat survey: The analysis of household food security in the semi-arid areas of Kenya

<i>Variables</i>	<i>Methods</i>	<i>Frequency of collection</i>
<i>Community-level</i>		
Food prices	Observe/recall	Once per visit
Non-food prices	Observe/recall	Once per visit
Population	Recall/records	Once per visit
Social services	Observe/recall	Once per visit
<i>Household-level</i>		
Income by source; farm, non-farm, etc.	Recall	Once per visit
Income by individual	Recall	Once per visit
Food expenditures	Recall	Once per visit
Non-food expenditures	Recall	Once per visit
Input by crop/livestock	Recall	Once per visit
Labour by crop/task, by household, by hire	Recall	Once per visit
Crop harvests	Recall	Once per visit
Crop and livestock sales	Recall	Once per visit
Education of head	Recall	Once at 1st visit
Age of head	Recall	Once at 1st visit
Leadership	Recall	Once per visit
Education of individuals	Recall	Once at 1st visit
Time allocation	Recall	Once per visit

DISCUSSION AND CONCLUSIONS

Data collection tools and procedures are well espoused in literature and knowledge about them may be regarded as commonplace. However, the collection of data for socio-economic research in the general field of agriculture in developing countries has its peculiar characteristics, limitations and challenges. If these challenges and limitations are not surmounted, and if the researcher cuts too many corners to avoid them, the results and conclusions derived from the intended research are bound not to make logical or academic sense.

The major and common tools used in collecting household data are the questionnaire and FGDs. Other approaches are the RRA, PRA, PLAR and PAR. Questionnaires, if carefully prepared and systematically administered, are likely to produce good data, germane to many socio-economic analyses. However, questionnaire surveys often take much time and resources and do not render themselves well to economic situations that obtain in developing countries. Poorly prepared and administered questionnaires will cause tendencies to amass superfluous data, adding to unnecessary costs and time wastage—and leading to generation of misleading results and conclusions. Questionnaire surveys also reduce the required contact between the researcher and the rural population. Adequate contact and interactions are necessary to ensure that 'outsiders' have sufficient grasp of situations and systems prevailing in the area of research interest.

Other surveys, that are more participatory, have much to contribute, especially when social and natural scientists combine their research efforts. They are usually cost-effective—an essential characteristic in poor countries—adaptable and open to unexpected information; allowing timely analysis, reporting, dissemination, and involving rural households as partners in research.

Participatory and mainly qualitative processes of data collection such as the FGDs and PRA may, nevertheless, be abused. They are often used as cheap and 'quick-and-dirty' means of collecting data, in the process compromising the quality and robustness of quantitative analyses that are often desirable in a number of situations.

We have noted that several impediments exist in data collection for agricultural economics and associated studies in developing countries. One of the most relevant in a developing country is

the uncertainty of future prices and costs. This is particularly a problem for relatively long-term studies, extending beyond a year.

NOTES

- 1 If one cross-sectional sample is considered over more than one time-period, it is referred to as panel sample, and the data so collected are called panel data (Nyariki, 2009).
- 2 There may be situations where a combination of two or more sampling methods discussed would be used.

REFERENCES

- Bouis H, Haddad L, Kennedy E 1992. Does it matter how we survey demand for food? Evidence from Kenya and the Philippines. *Food Policy*, 17(3): 349-360.
- Casley DJ, Lury DA 1987. *Data Collection in Developing Countries*. Oxford: Clarendon Press.
- Chambers R 1983. *Rural Development: Putting the Last First*. New York: Longman and John Wiley and Sons.
- Chambers R 1997. *Whose Reality Counts? Putting the First Last*. London: Intermediate Technology Publications.
- Devereux S, Hoddinott J 1992. Issues in data collection. In: S Devereux, J Hoddinott (Eds.): *Fieldwork in Developing Countries*. New York: Harvester and Wheatsheaf.
- Fowler FJ Jr 1984. *Survey research methods*. Beverly Hills: SAGE Publications, California.
- Freund JE, Williams FJ 1983. *Modern Business Statistics*. London: Pitman.
- Hoinville G, Jowell R 1978. *Survey Research Practice*. Hants: Gower, England.
- Mulwa FW, Ngululu SN 2003. *Participatory Monitoring and Evaluation*. Eldoret and Nairobi: Zapf Chancery and PREMESE-Olivex.
- Neefjes K 2000. *Environments and Livelihoods: Strategies for Sustainability*. London: Oxfam.
- Nyariki DM 1997. *Resource Availability and Productivity, Farmer Efficiency, and Household Food Security in Semi-Arid Kenya*. PhD Thesis, University of Reading, UK.
- Nyariki DM 2009. Price response of herd off-take under market liberalisation in a developing cattle sector: Panel analysis applied to Kenya's ranching. *Environment and Development Economics*, (Forthcoming).
- Odimegwu CO 2000. Methodological issues in the use of focus group discussion as a data collection tool. *Journal of Social Sciences*, 4(2-3): 207-212.
- Olsen W 1992. Random sampling and repeat surveys in south India. In: S Devereux, J Hoddinott (Eds.): *Fieldwork in Developing Countries*. New York: Harvester and Wheatsheaf.
- Scheaffer RL, Mendenhall W, Ott L 1979. *Elementary Survey Sampling*. 2nd Edn., Massachusetts: Duxbury Press, USA.
- Webb P, von Braun J 1994. *Famine and Food Security in Ethiopia: Lessons for Africa*. Chichester: John Wiley.