! TECHNICAL GUIDE #8 !

CONSTRUCTING SAMPLES FOR CHARACTERIZING HOUSEHOLD FOOD SECURITY AND FOR MONITORING AND EVALUATING FOOD SECURITY INTERVENTIONS: THEORETICAL CONCERNS AND PRACTICAL GUIDELINES

Calogero Carletto

IFPRI **International Food Policy Research Institute**
2033 K Street, N.W.
Washington, D.C. 20006 U.S.A.

March, 1999

CONTENTS

1. INTRODUCTION[1]

Reliable information on household food security is a prerequisite for the accurate and effective design, monitoring, and evaluation of development projects. In part due to the commitment, on the part of many development agencies, to work in marginalized areas, this information is often either not available or grossly out-of-date. But collecting data is not a costless exercise. This guide discusses how random sampling techniques—methods that use some mechanism involving chance to determine which farms, households, or individuals are to be studied—can economize on the costs of gathering information while increasing the likelihood that it will be both accurate and available in a timely fashion.

The guide has been divided into two parts: an overview and a series of technical appendices. The overview is written in a largely nontechnical fashion and is designed to be accessible to a wide audience. It begins with a brief explanation of why random sampling techniques are a powerful means of obtaining information on household characteristics such as food security. It then takes the reader through a step-by-step process of constructing a random sample. Having outlined these issues, a worked example is then presented. The second part consists of a number of technical appendixes that extend the discussion found in the overview. These are designed for individuals with some familiarity with statistics. The reader interested in pursuing the issues raised in this guide is encouraged to consult Bernard (1988), Casley and Lury (1987), Casley and Kumar (1988), Devereux and Hoddinott (1992), and Newbold (1988). More technical discussions are found in Kish (1965) and Cochran (1977).

2. WHY RANDOM SAMPLES?

Random Samples Rather than Censuses

One alternative to a random sample would be to obtain information on all observations, in a population census or a census of agriculture. The advantage of a census is that it would seem

[1] Funding for data collection and analysis of these data has been supported by the International Fund for Agricultural Development (TA Grant No. 301-IFPRI). We gratefully acknowledge this funding, but stress that ideas and opinions presented here are our responsibility and should, in no way, be attributed to IFAD.

to provide a very accurate "snapshot" of the population at a particular moment in time. It also ensures that groups that are numerically small (and hence possibly missed in a survey) are counted. Censuses are characterized by three characteristics: (1) individual enumeration (each unit of observation, say farm household, is measured separately); (2) universality within a defined territory or domain (information is obtained on every one in a certain area); and (3) simultaneity (everyone is interviewed at the same point in time). The key criteria is (3). The census should be conducted within a short and well-defined period of time to reduce omissions and duplications. However, there are a number of drawbacks to conducting a census.

First, it is usually much more expensive than conducting a survey. (This is not true, of course, where the population is very small.) Second, the processing and cleaning of a census is enormously time consuming. Further, a smaller sample allows the researcher to devote extra effort to ensure the information obtained is accurate, and the gains from having a smaller, more accurate survey could well outweigh the benefits of having less accurate information on a much larger group. Finally, many topics, such as those involving detailed transactions of individuals or firms require an intensity of interview or observation that cannot be carried out in the context of a census. So issues of cost, time, precision, and quantity of data all suggest that a survey is preferred to a census. There is a further reason. Censuses are unnecessary! We can learn all we need to know about a given population provided we draw a random sample of that population. This is referred to as inference. We draw a sample of a certain number of observations from a given population. We calculate parameters of interest such as means and proportions, which we infer represent the characteristics of the underlying population.

Random Versus Nonrandom Sampling

The discussion thus far indicates that it is not necessary to obtain information on all units of observation. Is it necessary, however, to choose those households or farms to be studied in a random, or probabilistic, fashion? Why not use nonrandom, or nonprobabilistic methods instead?

Nonprobabilistic methods are those in which the analyst consciously chooses who will be interviewed. Examples of these include the following. One is accidental sampling. This involves interviewing respondents as they are found, for example, walking down a track or road and interviewing whoever you happen to meet. A second is quota sampling. Here, enumerators are instructed to contact a specified numbers of observations possessing certain characteristics (e.g., 15 farms with no livestock; 10 farms with 1-3 head; 5 farms with more than 3 head of cattle). These quotas are assigned on the basis of what is known about the underlying population. However, the actual selection of observations is left up to the enumerator. A third method is purposive sampling. Here, individual units of study are chosen on the basis of some judgmental criteria. Suppose we wish to learn about long-term processes of environmental change in a rural area. To obtain this information, we could choose a sample of "wise old men." A fourth method is referred to as networking. Here, you find one person to interview and ask them to name others who are also suitable candidates, given the topic of interest. If the population is small, this can be a useful means of building up a sample. However, in larger populations every person or unit of observation does not have an equal chance of being sampled—that is, the sample selected is not representative of the underlying population. However, networked samples are very useful when exploring networks (you want to find people who know each other) or when dealing with hard-to-find groups.

Under a number of circumstance, nonprobabilistic sampling methods are appropriate. For example, if the population is homogeneous ("describe one unit of observation and you describe them all"), these methods produce information identical to that derived from probabilistic techniques. They are also appropriate if there is no intention to extrapolate the results to the larger population (for example, where the objective is to describe a village in general terms rather than obtain a statistically representative picture). Finally, such methods are useful where a sampling frame is unavailable or too costly to obtain.

But there are also significant drawbacks to these methods. Statements made on the basis of observations found in these ways must be limited to the sample itself—it is not possible to make legitimate inferences about the wider population. Further, standard statistical techniques—such as comparing means of two groups—cannot be used either. For these reasons, the use of nonprobabilistic methods by development practitioners is strongly discouraged. They should only be used where probabilistic methods are infeasible.

3. STEPS IN CONSTRUCTING A RANDOM SAMPLE

There are five steps involved in constructing a random sample: (1) determining the sample unit, (2) determining the "universe," (3) constructing a sampling frame, (4) deciding on the sample size, and (5) choosing the sample. Although these are discussed sequentially, it should be noted that it is often necessary to iterate back and forth among these. So, for example, practical considerations associated with choosing a sample may have an impact on the manner in which the sample frame is constructed and the calculation of the sample size.

Determining the Sampling Unit

The appropriate sampling unit—or unit of observation—is guided largely by the objectives of the survey and the project. For example, where a project seeks to increase farm yields, the relevant sampling unit for evaluation purposes would be the farm household. If the objective was to improve the nutritional status of children under five, the relevant unit of observation would be children in that age bracket. What is important here is that the definition of the sampling unit should be unambiguous and conform to local understanding and acceptance. The most common ultimate sampling unit in multipurpose socioeconomic studies is the household, even if individual-specific estimates are sought. In some countries, there may be a generally accepted definition of what constitutes a household—for example, the definition adopted by the Central Statistical Office. Even where such a definition exists, it should be validated locally before proceeding with the listing exercise.

Determining the Universe

The "universe" is the location or population or group that the study seeks to describe. Again, this is likely to be determined by the objectives of the project. If the project is located in, say, western Honduras, then western Honduras would be the location of the study. However, it is not always practicable to survey the entire location. The discussion below on "choosing the sample" and the worked example from Malawi illustrate solutions that are available when this problem arises.

Constructing a Sample Frame

The use of probabilistic methods to select a sample requires a sample frame:

> The frame for a sample is a list of the units in the population (or universe) from which the units that will be enumerated in the sample area are selected. It may be an actual list, a set of index cards, a map, or data stored in a computer. The frame is a set of physical materials (census statistics, maps, lists, directories, records) that enables us to take hold of the universe piece by piece (Casley and Lury 1987, 52).

Examples of lists that can be used as sample frames include lists of administrative areas, census materials, ordinance survey maps, tax listings, land registries, and lists of project beneficiaries. In practice, there are a number of dangers when working with such materials. Take, for example, a list of households. First, the frame may be inaccurate. This could result from errors in recording information—names might have been misspelled, adults were listed as children, households contain more people than is recorded, and so on. Alternatively, these might have occurred because the information was collected from neighbors as household members were absent or unavailable when the frame was created. Second, the frame might be incomplete. Households or groups of households may have been omitted This might have occurred because the frame is out-of-date (for example, households have subdivided or migrated in or out) or because of poor enumeration when the frame was created. There might have been difficulties in determining the location of boundaries, with the result that certain households were missed. Third, there might be duplication. Some households are included twice, possibly because (1) the lists were compiled by more than one person; (2) confusion over names; or (3) disputes over land claims. Devereux (1992) provides a good example of some of these problems in his description of surveying households in northern Ghana.

> When I first arrived in my chosen village of Pusiga I introduced myself to the sub-chiefs in the two sections, Terago and Tesnatinga (or Teshie), in which I planned to work. These sub-chiefs had recently compiled lists of households for their sections, which were used by the District Administration to distribute small quantities of government food aid (following the two successive poor harvests mentioned above). Had these

lists been compiled for an unpopular purpose such as tax collection, I would have had reservations about their accuracy. But since everybody had an incentive to register for food aid, I decided to use the sub-chiefs' lists as a basis for household enumeration.

Nonetheless, these lists were inaccurate in several respects. . . Over-reporting occurred mainly in large, complex compounds, and typically took the form of young men claiming to be separate households when they were, in fact, still farming with their brothers or father. The explanation for this was simple. When the household lists had been drawn up, local residents were well aware that the purpose was to distribute food aid. People in large compounds reasoned that if each household was to receive free food, it was to their advantage to exaggerate the number of separate households in their compound. When I made my first round of interviews, the expectation that I would be bringing some kind of free or subsidised assistance to the village was high, and over-reporting was standard practice. During the year I gradually discovered which compounds had over-reported household numbers, and simply crossed them off my list. (A clear indicator was when I asked several 'household heads' in a compound about planting, harvests, and asset ownership, and received identical or near-identical figures—since they were each listing, in fact, the same (joint) production and assets.) . . .

Under-reporting of households occurred most commonly with old women, especially widows. Although most old widows are looked after by a son or son-in-law, this is not always the case, and some old women constitute separate households, either because they insist on retaining their independence by farming their own land, or because they have been cast adrift to fend for themselves. In my sampling frame there were three such single-person 'households', one in the first category and two in the second, all of which I missed until it was too late to incorporate them in the lists of households from which my samples were randomly selected.

> The reason why these widows were missed is to be found in the local conceptualisation of a household, which corresponds broadly to the Western notion of a 'production-consumption unit'. A man is said to constitute a separate household if he is 'farming separately' (from his father and brothers) and 'feeding himself' (and his wives and children)—that is, both a production entity and a consumption entity. The two widows living on their own were virtually beggars, being too infirm to work and having no-one to help them with farming. In fact, they were dependent on handouts from relatives and neighbours. So they did not strictly qualify as households in terms of the local definition because they were neither 'farming separately' nor 'feeding themselves'.

It follows that survey designers should always plan to have any existing list checked. In monitoring and evaluation exercises, the population under study, or at least a domain of it, is generally composed by the beneficiaries of a certain project or program. In many instances, the lists of project beneficiaries is readily available with the project management. However, even in these apparently favorable conditions, it is imperative to check these lists for inconsistencies, omissions, and duplications. By no means should its accuracy be taken for granted.

Where no such listing of households, or units of observation, is available, or where such lists are so outdated or inaccurate as to be useless, two possibilities remain. These are (1) to create a list; or (2) to derive a sample without a frame. These are discussed in turn.

Creating a sample frame can be a time consuming and expensive exercise. For this reason, there may be practical advantages associated with using multistage sampling (described below) and/or restricting the "universe" to be studied. For example, an evaluation of a project might be limited to certain localities rather than all areas in which a project has operated. It is important to note that by adopting such strategies, probabilistic samples are representative of a *restricted* universe and as such any extrapolation of the results should be confined to it.

One approach is to start with a list, even one that is known to be inaccurate. For example, in northern Mali, survey work began with lists of households that had been compiled several years previously for the distribution of food aid. The survey team, accompanied by village leaders, walked through the villages matching names on the list to households, adding new names, and deleting those no longer resident in the village.

Where even rudimentary lists are unavailable, maps can be used as a starting point. A first step in area sampling may involve the use of a map providing a graphical representation of the universe, e.g., a region or a province. Using easily identifiable natural boundaries, the map can then be partitioned into approximately equal sized segments. Once all the segments/villages have been delimited and some chosen, sketch maps can be easily produced in a relatively short time without need of much expertise. Of course, the amount of time and resources going into this mapping exercise should be suited to the objective at hand. In most cases, very rough sketches describing the main roads and pathways and some landmarks (e.g., a church, a mosque, a borehole, a river, etc.) clearly delimiting different sub-areas of the segment/village could suffice. In most cases, however, the inclusion in the sketch of the individual domiciles, properly numbered, may be necessary. As in the case of working with listings, it is important to verify that no area or sampling unit of the universe have been missed and that no overlapping occurs between different maps since this would obviously result in unequal probability of selection for the elements of the chosen population.

But there may be instances where it is simply infeasible to construct a sampling frame. In such circumstances, the following two-stage technique—EPI Cluster Survey Design, developed originally to monitor and evaluate the Expanded Programme on Immunization (EPI)—can be used.

The original design, used for the monitoring of immunization coverage of children within a target age (generally 12-23 months) involves the selection of 30 primary sampling units or clusters (villages or other types of area units), and the subsequent drawing of 7 ultimate units (children) from each cluster, for a total sample size of 210 children. The clusters are selected from a comprehensive list of villages or area units with probability proportional to estimated cluster size (see Appendix 4). Census information and administrative records may be used to generate the list containing the estimated size of the cluster. The second-stage selection of 7 children in each cluster was originally envisioned as a random selection from a list of children in the target age living within the cluster. However, difficulties in the enumeration has given origin to more simplified procedures. A commonly used variant of the original scheme contemplates choosing a random direction from a central point in the village/area unit by spinning a pen or bottle. All, and only, the households along this direction up to the edge of the cluster are enumerated and one chosen at random. Starting from the chosen households, and along the direction line, seven adjacent households with children in the target age are selected and interviewed.

A plethora of variants have loomed in recent years to partly overcome some of the limitations associated with the standard design. While choosing seven adjacent households in the case of restricted target group, i.e., children between 12-23 months, may actually result in a quite spread out sample within the cluster, in another circumstance in which eligibility criteria are not so stringent, the selection is likely to identify a highly concentrated conglomerate of households. Under the plausible assumption that adjacent households exhibit very similar socioeconomic characteristics, it appears evident why the standard design does not perform well in multi-indicator socioeconomic surveys.

With the goal of selecting more heterogeneous elements within the cluster, a possible variant to the standard design would be to select the third, fourth, or fifth households starting from a central (or randomly chosen) location after a direction is picked. From this last selected household, we repeat the procedure and proceed in a random walk fashion until the quota is met. Alternatively, the village could be split into smaller areas and from the center of each unit (or any randomly chosen point), a direction picked and the *N*th household meeting the eligibility criteria interviewed.

Deciding on the Sample Size

Calculating sample sizes is one of the most technically demanding aspects of survey design. Although a number of software packages—such as *Epi-Info* and *STATA*—automate these calculations, it is still necessary to understand what information is required in order to run these routines. This subsection provides an overview of these issues. A more thorough, but more technically demanding discussion is found in Appendix 2.

Abstracting from practical issues such as the time and resources available to undertake a survey, a decision regarding sample size is strongly linked to the required level of precision in the variables we seek to measure. Precision—or sampling error—is described in terms of a margin of error and a confidence level. For example, we might wish to estimate sample maize yields within three percentage points of the true mean (the mean we would obtain if we measured all maize yields). This statement implies that if we were to take 100 samples, we would expect that the sample means would be within three percentage points of the true mean at least 95 times. Three other factors will also play a role. One is the distribution of the variable of interest. If maize yields are identical across all households, then we would only need to sample one household in order to determine the average level of maize yields. By contrast, more dispersed distributions require a larger sample size. Second, sample sizes are affected by the particular sampling design chosen. Multistage designs require larger samples than single stage designs in order to achieve the same degree of precision. Third, increasing the number of variables that we seek to estimate may also affect the sample size needed to attain a certain degree of precision.[2]

Finally, we note that the belief that sample size depends on the size of the population and therefore on the sampling fraction is widely spread but erroneous. The size of the population affects only marginally the precision of the estimate. The precision of the estimate is directly related to the absolute size of the sample, but much less so to the sampling fraction. A sample of 100 units drawn from a population of 1,000 (sampling fraction 10 percent) is highly unlikely to produce more precise estimates than a sample of 200 from a population of 10,000 (sampling fraction 2 percent).

Choosing the Sample

Armed with a suitable sample frame that lists our units of study and knowledge regarding the desired sample size, the last step is to select the sample in a random or probabilistic fashion. There are four types of probabilistic methods: systematic, simple random sampling, stratified, and multistage.

A relatively straightforward method of selection is systematic sampling, where draws are made at fixed intervals through the list starting from a random unit. For example, suppose we wish to extract the same sample of 10 households from the list of 150 households. We randomly select a number between 1 and 15 (150 divided by 10) and, starting from that unit, we select

---

[2] Please refer to Appendix 2 for a more technical presentation of these factors

every 15th household. If 5 were the randomly selected number, then our sample will be composed of households 5, 20, 35, 50, 65, 80, 95, 110, 125, and 140.

Note that in addition to being a random selection method, this method has another advantage when the list is ordered on the basis of some feature related to the variable of interest. Suppose we wish to estimate crop yield and the list is ordered based on farm size class, then systematic selection would guarantee that a wider spectrum of farm size classes are represented in the sample. Following this systematic method, we can be almost certain that the first sample element (household 5) belongs to a different class of, say, element 8 (household 110) or 10 (household 140). Set against this advantage is a potential danger. If there is some subtle, difficult-to-observe ordering of the sample (resulting, for example, in small farms never having numbers ending in zero or five), the observations drawn will not be a random sampling of the population.

A second method is systematic random sampling. A simple illustration of this is the following. Write all the farm identification numbers on individual slips of paper and throw these in a hat. Shake the hat vigorously. Pick out the number of farms you want to interview and that is your sample. However, in large populations this is a rather tedious operation (and might require a very large hat!). An alternative method is to use a table of random numbers. Appendix 3 explains how these are used.

There is a potential weakness with both systematic and systematic random sampling. Suppose we are drawing a sample of 100 farms from a population of 1,000. We know from the census that 30 percent of these have more than 10 acres of land, so our sample should contain 30 such farms. However, this is only true on average! Though the likelihood is high that our sample will contain 30 large farms, it is also possible that it contain 20, 25, or 40. Suppose it only contains 15 such farms. If larger farms have better access to formal sector credit than smaller farms, ceteris paribus, and given that the larger farms have been underrepresented, you might feel that inferences regarding credit will not be reliable. One tempting possibility would be to pick two or three other samples and choose the one you thought was most representative. The difficulty with this approach is that the sampling procedure being used—the population is sampled until you find a sample you like—can no longer be justified and the results are no longer suitable for the purposes of inference.

There is a solution to problems such as these: random stratified sampling. The first step is to divide the population into groups or strata. Here, the division would be between the 300 large farms and the 700 smaller ones. Using the random number method, select 10 percent of farms in each category, so the resultant sample contains 30 large farms and 70 small ones. The proportions in the sample are identical to those in the underlying population.

Random stratified sampling is an attractive means of obtaining a sample. However, it is helpful to note two potential problems. First, the relevant stratification variables must be known in advance. Second, you must know the underlying population proportions of each strata. Addressing these problems requires additional information on each unit of observation. For example, lists of farms may only contain the name of the household head. A short survey may be necessary to obtain information to stratify and this may be too time-consuming or expensive.

A fourth form of sampling is multistage or cluster sampling. Whenever the universe from which we wish to draw the sample is geographically spread out, single-stage procedures such as SRS or systematic sampling may not be logistically feasible since they are both likely to generate equally dispersed samples. The necessity to lower transportation and organizational costs, as well as reduce nonsampling errors (enumerators working on a large area may be more difficult to supervise, increasing the likelihood for errors) suggest that a multistage design may be more appropriate. In addition, multistage designs can produce substantial savings in terms of time and financial resources that must be allocated to the listing operations.

A two-stage design would generally call for the selection of geographically-delimited nonoverlapping primary sampling units (PSU), also known as clusters (examples of clusters are a region, a district, a village), the selection of a limited number of clusters, and within each cluster, the random selection of a certain number of ultimate sampling units. Given that a two-stage design is chosen, a number of issues arise. How do we select the clusters from the universe? How many clusters do we select? How many ultimate units do we draw from each cluster?

The way clusters are selected depends primarily on the availability and accuracy of a complete sample frame. In the simplest case scenario in which such a list is available and the clusters are of equal size, we can select a number of them using simple random sampling and, within each, draw an equal number of ultimate units. A complication arises when clusters are of unequal size. This is generally the norm. In this case, the survey planner has several options, which are described in Appendix 4.

4. A WORKED EXAMPLE

This example outlines how a random sample of farmers was obtained in order to assess the impact of two projects directed toward smallholders in Malawi. As discussed in Technical Guide #10 (Evaluation Methods for the Promotion of Household Food Security in Rural Development Projects), it was necessary to survey participants in both projects, as well as households enrolled in neither ("control households"). The example illustrates practical difficulties encountered in sampling, the solutions adopted, as well as the time requirements of the different steps.

Selecting the Sampling Unit

Both projects target smallholder farmers. Consequently, the sampling unit is a smallholder farm household, classified using a local definition as a rural household with less than 10 hectares of land.

Selecting the Universe

The next step was to select the area(s) for the data collection. Based on a classification by the Ministry of Agriculture, the country of Malawi is divided into three regions (North, Central, and South), further divided into Extension Planning Areas (EPAs). Although it would have been ideal to work in all three regions, time and budgetary constraints made it necessary to restrict the survey to a single region. Field visits conducted in the regions (one or two days for each visit, combined with extensive talks with key informants) revealed significant differences between these regions. For this reason, a random selection of one region was not appropriate. Instead, it was decided to select an EPA in Central region. To facilitate the contrast of the two projects and rule out differences in location-specific features (or having to control for them during data analysis), it was decided to select an EPA in which both projects were active. This restricted the choices to a pool of only two EPAs with very similar characteristics: one EPA was randomly selected. An implication of this decision was that it was not possible to extrapolate any findings from this region to the whole country.

Constructing the Sampling Frame and Selecting the Sample

Given that the objective of the study was to compare the two projects against each other, and against the control group, it was necessary to sample households in both projects as well as households in neither project. These three groups constitute separate "domains." (Technically, the "universe"—the EPA—was stratified by domain.) One way of doing so would have been to enumerate all households in this area and select households from each domain in proportion to their number in the EPA. However, given the relatively low coverage of both projects, this technique would have led to the selection of an insufficient number of observations among the two beneficiary groups. Since the main objective of the study was the contrast of these groups and not the extrapolation of the group or domain estimates to the EPA as a whole, we chose to select an approximately equal number of observations in each of the three strata, namely the food security and agriculture development project beneficiary groups, and the control group.

Smallholder farmers belonging to either project are organized into clubs of variable size between 10 and 30 households. The club was selected as the primary sample unit. The lists of clubs belonging to each project were available with the project management units. But because membership in these groups changes radically over relatively short periods of time, these lists were not considered reliable. Further investigation (one or two days talking to key informants) revealed the existence of several such lists. In some instances, a list would differ from the others quite substantially. We spent several days trying to reconcile the different sources in an attempt to come up with a unique list that reflected actual project membership.

The first step in the verification process was to clearly define membership for each project. Given the objective of the stratification (to enhance the group contrast and measure project impact within each domain), a club was considered a beneficiary of project A if it had been active within the project for at least two seasons and it had never belonged to project B. By active we meant that it had produced and sold tobacco in both seasons and had participated in the project’s activities. Based on the definition, some clubs were excluded from the list either because of dual membership or because they had not produced and sold any tobacco.

In the case of the food security project, determining membership was slightly easier since it could be related to access to the credit package being disbursed by the project. A club was considered an food security beneficiary if it had received a full or partial credit package in both of the last two seasons. The main difficulty with this group was represented by the common practice for members to use different names in joining the club. Local key informants—field assistants and village headman—helped screen "hidden" duplications.

Once we accounted for these sources of omissions and duplications, we ended up with a list of 14 clubs for the food security project and 71 for the agricultural development project. The next step was to enumerate all the members of each club. These lists were not available. The only information readily accessible was the total number of members at the year of club formation. Given the dynamic nature of membership, these figures were not considered reliable. Therefore, enumeration of the clubs was deemed necessary. To reduce the amount of time necessary for this operation, it was decided to select only a limited number of clubs from the agricultural development list. To this end, 30 agricultural development clubs were selected, using a fixed probability of selection. (This two-stage EPSEM design was chosen because the absence of accurate estimation of the current cluster sizes meant that PPS methods would have required the use of weights at a later stage. See Appendix 4 for further explanation.)

Once enumeration for these 30 clubs had been completed, because of the variable size of the clusters, we drew from each cluster a number of households proportional to the size of the cluster. The procedure resulted in a self-weighting sample within the agriculture development domain. Due to the already limited numbers of food security clubs eligible for inclusion in the sample, a full census was considered appropriate for this domain.

The selection of the control group called for a different methodology altogether. Available census data were more than 10 years old, and hence suspect. Alternative administrative records were not available. Tight time constraints made complete enumeration of the selected villages infeasible. In addition, to reduce transportation costs and to avoid selecting households from villages where neither project was active, it was decided to select a control household for every other beneficiary household in the village in which this latter resided. One complication was that the exact village of residence of the beneficiary household was not known until the household was actually visited, so enumeration in advance of selected villages was not possible. In addition, time constraints would not have allowed for it. For the selection of the control households, a variant of the EPI cluster design was used.

Once we visited a village in which selected beneficiary households lived, we randomly selected one nonbeneficiary household for every other beneficiary household in the sample. As a way of explanation, we give an example. Assume that a total of eight beneficiary households belong to village *X*. A total of four nonbeneficiary households were to be drawn from this village. The first step was to roughly sketch the village to locate a central point. From this central point, a team of enumerators, jointly with a supervisor, chose a random direction by spinning a pen on a flat surface. Once a direction was selected, the enumerators were asked to follow that direction and starting from the 4th dwelling, interview the first household that met the eligibility criteria to belong to the group, i.e., they owned less than 10 hectares of land and they had never belonged to either the food security or the agriculture development project. If the same team had been assigned another control household, the supervisor would again spin the pen in front of this first selected household, choose a new direction and, starting from the 4th dwelling, identify the next household to be interviewed along this "random walk." If instead, another team was required to select an additional control household in the same village, the random walk would start again from the center of the village by randomly choosing a new direction.

One of the potential problems with this variant of this selection design is that it tends to underrepresent households located in more remote areas from the village center. To partly obviate to this problem, bigger villages were often divided into sub-areas and a center chosen in the sub-area in which the selected beneficiary household fell. Another potential problem was that the selection must follow a natural path, restricting the number of options in terms of the direction an enumerator can take from the central point. Whenever possible, the enumerators were instructed to cut through fields and follow as closely as possible the direction chosen.

Calculating the Sample Size

Sample size calculations took into account the degree of precision required, statistical power, design effects, and estimated nonresponse rates. (These are fully explained in Appendix 2.) A total of 202 households per stratum, for a total sample size of approximately 600 households, was pursued. The value corresponds to an prevalence of about 0.7, a difference in the magnitude of 0.15, a one-tailed statistical significance of 95 percent, a statistical power of 80 percent, and a design effect of 2. Plugging in the values in equation (3) (found in Appendix 2), we obtained

$$2 \cdot [(1.64 + 0.84)^2 \cdot (0.21 \cdot 0.1275) / (0.15)^2] = 184 .$$

Considering a nonresponse rate of 10 percent, we obtain a total of 202 households per stratum.

APPENDIX 1

A GLOSSARY OF SAMPLING TERMS

universe	the location or population or group that the analysis seeks to describe
sampling units	the unit of observation of the study: farms, households, individuals, etc.
sampling frame	the list of sampling units. It must contain all units within the universe.
self-weighting samples	a sample in which all units have an equal probability of selection
sampling fractions	the ratio between the sample size and the population size. Also called the *selection probability*
domain	a part of the population for which separate estimates are sought. Examples are farms of a certain size, individuals of a particular age group.
cluster	the aggregation of sampling units, often based on geographic proximity. Examples are a village or a section of village.
take	the number of sampling units drawn from a selected cluster

APPENDIX 2
CALCULATING SAMPLE SIZES

The Basics

What is the required sample size given a minimum required level of precision of the estimate? The level of precision sought can only be achieved within an acceptable margin of error (an absolute percentage) and with a given level of confidence. In practice, the confidence level, as expressed in percentage terms, must be converted into what is termed the normal deviate, indicating the number of standard deviations defining the confidence interval. Table 1 reports the most commonly used conversion values from confidence interval to normal deviate. Attention should be paid on whether we are interested in a one-sided or a two-sided interval.

Table 1 Normal deviates for confidence intervals

Confidence level (percent)		**Normal deviate**
Two-sided interval	**One-sided interval**	**K**
80	90	1.28
90	95	1.64
95	97.5	1.96
98	99	2.33
99	99.5	2.58

Example: Assume that we wish to estimate the prevalence of chronic malnutrition. Specifically, we may wish to get estimates with a margin of error of 5 percentage point, plus or minus, with a confidence level of 95 percent. Assuming a prevalence of malnutrition in the universe of 40 percent, then 95 times out of 100, the sample estimate will fall between 35 and 45 percent. To compute the required sample size, we would use the value in the third row of the two-sided confidence interval, 1.96. If, instead, we only wish to ensure that the estimate will not exceed the population value by more than 5 percent, then we would pick the one-sided value at 95 percent, 1.64.

As summarized in the formula below, the practical requirements for the calculation of the sample size are (1) the variability of variable within the population, (2) the acceptable margin of error of the estimate, (3) the degree of confidence, and (4) the sample design:

$$n = d \cdot K^2 \cdot \sigma^2 / M^2 , \qquad (1)$$

where n is the required sample size, d is the design effect, K is the normal deviate, σ is the estimated variance of the indicator in the population, and M is the allowed margin of error. d, the design effect, is defined as the ratio between the variance associated with a multistage design and the variance, had the sample been drawn using SRS. (This is explained later in this appendix.)

Example: For the moment, ignore the design effect and assume a SRS sample (for which the design effect equals unity). Imagine we wish to estimate the percentage of infants between the age of 0 and 12 months with a complete immunization record. Assume that previous studies in the area (or similar area) had estimated a prevalence (p) of 30 percent and that we wish to estimate the value with a margin or error of +- 5 percent with a 95 percent confidence level. First we would need to convert the set confidence level into the corresponding value of the normal deviate. By looking at the table, the value of the normal deviate for a two-sided interval at 95 percent confidence level equals 1.96. For an assumed immunization coverage in the universe of 30 percent, then = $p(1 - p) = 0.3 * 0.7 = 0.21$. Substituting $M = 0.05$ (5 percentage points), it follows that $n = 323$.

Table 2 reports the required sample size for some commonly used combinations of the parameters P and M, assuming a standard 95 percent confidence level and no design effect.

Table 2 Sample sizes, given different prevalences and margins of error

	Margin of error (M)		
Prevalence (P)	**3%**	**5%**	**10%**
10%	385	138	37
20%	684	246	62
30%	896	323	81
40%	1,024	369	92
50% (upper bound)	1,066	384	96

For values of P greater than 50 percent, use the value of P that differs from 50 percent by the same amount. For example, for P = 70 percent, we would use the values corresponding to P = 20 percent. This is valid because of the symmetry property of the normal distribution.

Note that in order to double the precision of the estimate, we must quadruple the sample size. It is also obvious that the sample size is more sensitive to changes in the precision requirements than in the prevalence. For this reason, if no information is available about the variability of the variable in the population, the most conservative value of 50 percent (in the case of a ratio estimate) is generally adopted without major effects on the sample size.

If separate estimates are required for distinct domains, then the total sample size will equal the sum of the observations required to achieve the targeted precision for the specific variable within each domain. For equal size domains, and under the assumption of equal variance of the indicator in both domains, the sample size should be double the single domain requirements.

Adjusting for Multi-Indicator Surveys

In the vast majority of surveys, we are interested in estimating the values of multiple indicators. In these circumstances, it is unlikely that the required level of precision for the different indicators, as well as their prevalence within the population, will be equal. The required sample size will be based on the most demanding estimate. For example, let us assume that, as part of the same survey, in addition to immunization coverage, we wish to estimate another variable, e.g., the level of chronic malnutrition as measured by the percentage of stunted children between the age of 6 and 60 months, with a margin or error of 3 percent at 95 percent confidence level. We also have information that the approximate prevalence in the universe is 40 percent. The sample size to meet these requirements will be 1,024. The required sample size of the multi-indicator survey will now have to be 1,024 (no longer 323) if we want to guarantee meeting the precision targets for both variables.

Adjusting for Different Sampling Units

Although many indicators of socioeconomic surveys refer to the individual household member, the ultimate sampling unit of choice is more often the household. This has implications on the computation of the sample size. Following up on the above example, assume that available information suggests that each household in the area under study has, on average, 0.4 infants between the age of 0 and 12 months, and 1.8 children between the age of 6 and 60 months. Assuming that in each household visited, the enumerators are asked to take the anthropometric measurements of all children age 6-60 months, then in order to reach the desired number of 1,024 children, we will only need to visit about 569 households to meet the precision requirements for this variable (on average, we would collect information on 1.8 children per household). Conversely, in order to have information on the immunization records of 323 infants age 0-12 months, the enumerators may have to interview 808 households (323/0.4). It this is the case, the required sample size to guarantee precision of both indicators would now be 808 households (note that the most demanding indicator is reversed with respect to the previous situation).

Design Effects in Multistage Surveys

A concept used for the assessment of how cluster design affects precision is given by the design effect (deff), defined as the ratio between the variance associated with the multistage design, and the variance if the sample had been drawn using SRS. This can be expressed as

$$d = 1 + \rho(m - 1) \,, \qquad (2)$$

where ρ is the intraclass correlation and m is average number of responses in the cluster. For example, a value of $d = 1.5$ implies that, to achieve the same level of precision, the sample size of the two-stage design would need to be 50 percent larger than its SRS counterpart. Therefore, if the cost function is such as to allow for an increase in the sample size of more than 50 percent, a two-stage design is more feasible, since at equal cost we may achieve higher precision of the estimates. Note that, for a given intra-cluster correlation, the only way to contain the design effect is to keep the average "take" low.

Example: Continuing with the previous example, in which a sample of 323 units was required using SRS, let us now assume that we wish to achieve the same level of precision using a two-stage cluster design and that 7 units per cluster were to be drawn. Assume that the intra-cluster correlation for the variable of interest equals 0.2. Under these assumptions, we are able to quantify the design effect, which will equal 2.2, implying that a sample of 711 would be needed to estimate the same variable with an equivalent level of precision, i.e., we would need to select more than 100 clusters.

A practical problem in computing the design effect is the need to quantify intra-cluster correlation of the responses. As a rule of thumb, generally a value between 0.1 and 0.3 is assumed if more detailed information of the intra- and inter-cluster variability is not available. For example, assuming a value for the intra-cluster correlation of 0.15 and a take of seven elements in each cluster, then the design effect would equal 1.9, indicating that to achieve the same level of precision of SRS, the size of the two-stage sample must be almost double.

Generally, a design effect of 2 is used as a standard. While appropriate in most instances, the survey planner must make sure that he is not dealing with special cases in which the correlation of responses within clusters with respect to the variable of interest is unusually high. Examples of that could be location-specific questions that are likely to produce strongly correlated responses among elements in the same geographically delimited cluster, such as the use of health facilities, availability of water, etc.., for which the value of intra-cluster correlation must be set to higher levels, say 0.4 or more.

The final issue relates to the choice of an optimal take. Earlier we had posed the question on how many clusters we should select and how many units per cluster, for a given sample size. In the multistage case, the sample size is no longer given, since it will depend on the particular design chosen. Also, setting up two parameters will suffice, since the third will automatically follow. Then, which two parameters do we set up, or better, what is the sequence to be followed in choosing the parameters?

Since the sample size will now depend on the particular cluster design chosen, as quantified by the design effect, based on the formula reported above, this effect can be computed in terms of the size of the cluster. How do we decide, then, on an optimal take of the cluster, which, in turn, will determine the required sample size, and the necessary number of clusters? Unfortunately, no golden rule exists to decide on an optimal take. Generally, considerations over the cost of the different combinations, and the implications on sampling error, will guide the decision. However, the difficulties associated with obtaining accurate information on the cost function gives way to the application of rules of thumb. One such basic rule of thumb often suggested for multipurpose socioeconomic surveys is to keep the take within single digits, whenever possible. With this in mind, a score of practical considerations will also play a role in the decision.

For example, the viability to provide a full daily workload to each enumerator may influence the choice on the take within each cluster. Assume that the field team is composed of three enumerators and that, based on the length of the questionnaire and the spread of the clusters, we guestimate that about three interviews per day can be conducted by each enumerator. A take of 9 would be recommendable in such a situation, since it would avoid keeping some of the enumerators idle for part of the day or splitting the team in two different clusters during the same day.

In summary, the choice of a particular cluster design has obvious implications in terms of lost efficiency. This loss can be quantified in terms of design effect, which measures the relative inefficiency of multistage design with respect to a benchmark, SRS. The losses associated with cluster design must be weighted against the cost gains due to simplified logistic arrangements, and will depend on the particular cost function characterizing the survey. Both theoretical and practical considerations will enter the decision about the optimal take of the cluster to achieve a target precision requirement with a multistage design, within budgetary constraints and field conditions.

Calculating Sample Sizes for Project Monitoring and Evaluation

The monitoring and evaluation of a project generally involves the evaluation of changes over time or the comparison of two groups. The indicator to be measured is generally expressed as a mean, a proportion, or a total value. For example, one may wish to contrast differences in yields across project beneficiaries and nonbeneficiaries, or the adoption rate of a certain technology between these two groups. Or, alternatively, the goal may be to monitor the changes over time in the adoption rate of a certain technology since the project onset. What we ultimately wish to accomplish in this type of monitoring and evaluation exercises is the assessment, with a certain degree of confidence, that a change/difference is detected, if it occurred.

The design of a survey for this purposes involves some modifications to the basic calculations presented above. Sample size calculations require information on the following factors: the number of units in the population; the baseline level of the indicator (if over time) or the level for control group; the magnitude of the change (if over time) or the difference; the level of statistical significance, i.e., the degree of confidence in assessing that the change/difference observed did not occur by chance; and the level of statistical power, i.e., the degree of confidence that the change/difference will be detected if occurred.

For the purpose of project evaluation, the concept of statistical power is perhaps most important. Lack of power is likely to lead us to conclude that no difference exists between the two groups when, in fact, actual changes have occurred but were not detected because of insufficient sample size. To that end, the survey planner wants to ensure that a sufficient number of observations are selected so that some minimum power requirement is met. A value of 0.8 is generally adopted as standard, although statistical power in the range of 0.9 should be sought whenever possible. The latter would ensure that if a difference exists, it will be detected 90 percent of the time.

Conversely, the concept of statistical significance relates to the level of confidence of concluding that change has occurred when it has not. The adopted standard of 95 percent assures that we would incur in the risk of falsely reporting a change/difference no more than 5 times every 100 samples.

If the indicator for which we wish to estimate the change is a proportion, the formula to compute the sample size in each group is given by

$$n = d\,[\,(Z_\alpha + Z_\beta)^2 \cdot (P_1(1 - P_1) + P_2(1 - P_2)) \,/\, (P_2 - P_1)^2\,]\,, \qquad (3)$$

where n is the required sample size per group or survey round, d is the design effect, P_1 and P_2 are the level of the indicator in each group or period, Z_α and Z_β are the normal deviate values corresponding to level of statistical significance and statistical power, respectively. As in the single group case, we need to convert the level of confidence into normal deviates. For the statistical significance parameter, the one-sided column of Table 1 can be used. For statistical power, the following values can be applied.

Power (%)	80	90	95	97.5	99.9
	0.84	1.28	1.64	1.96	2.32

For reference, we also report the required sample size for the given combinations of P_1 and the magnitude of change to be detected (P_2 - P_1) for α = 95 percent, β = 80 percent and d = 2.

	Magnitude of change/difference (P_2 - P_1)					
P_1	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
0.1	1,075	309	152	93	63	45
0.15	1,420	389	185	110	73	52
0.2	1,176	457	213	124	81	56
0.25	1,964	513	235	134	57	60
0.3	2,161	556	251	142	90	62
0.35	2,310	587	262	147	92	62
0.4	2,408	606	268	148	92	62
0.45	2,458	611	268	147	90	60
0.5	2,458	606	262	142	87	56

Source: Sampling Guide, Food Security and Nutrition Monitoring (IMPACT) project.

If the indicator is, instead, a value such as a mean or a total, the formula for the sample size is

$$n = d\ [\ (\ Z_\alpha + Z_\beta)^2 \cdot (\sigma_1^{\ 2} + \sigma_2^{\ 2}) \ /\ (X_2 - X_1)^2\ ]\ , \qquad (4)$$

where σ is the deviation and X_1 and X_2 are the level or mean values of the indicator for these groups or periods.

Note that this calculation requires knowledge of the variability of the indicator. The use of values from previous surveys or from studies in similar contexts is recommended. If not available, the values must be guessed. Consulting an expert statistician with knowledge of the area and the topic at hand may be appropriate.

Repetition Versus Rotation

Monitoring exercises usually involves the periodic collection of information among the program participants. This repetition of the survey operations raises a number of issues that we briefly address in this section. First is the issue of whether the same group should be used year after year, or replaced, entirely or in part. The answer will depend on the objective of the study, as well as a number of other considerations, such as the possibility for contamination or the respondents level of cooperation. In general, the use of the same sample over different years may be more appropriate for the monitoring of project impact, while renewal of the sample each time may be more suitable when the objective of the study is simply to estimate a certain parameter. However, there are exceptions to this basic rule dictated by more practical considerations. For example, the perception of an increasing unwillingness by the respondents to participate in the survey may be a good enough reason to renew, totally or partially, the original sample. Also, the increased likelihood for contamination in the responses as time goes by may also prompt a partial or total replacement of the sample.

Finally, under the assumption that maintaining the same sample may be preferable in certain circumstances, when calculating the sample size for a monitoring exercise involving the repeated collection of data over several years should also take into account the likelihood that these same households can be interviewed in different survey rounds. To the problem mentioned above related to the increased unwillingness by a respondent, we must also account for natural phenomena such as the likelihood of death, illness, or migration of the respondent, and household dissolution. A larger sample size than the precision requirements call for may be more appropriate in the baseline survey if we wish to ensure the same precision year after year.

APPENDIX 3

USING RANDOM NUMBER TABLES TO OBTAIN A SAMPLE

These are found in most statistics textbooks (the example below is taken from Bernard (1988), pages 460-461) or can be generated using standard statistical software packages or even, in some cases, a pocket calculator.

Suppose from our population of 1,000 farms, we require a random sample of 100. We take some arbitrary place in the table, say the sixth column, and read off the first three digits. The first five numbers are 548, 240, 025, 311, and 366 and so the farms with numbers 548, 240, 025, 311, and 366 would appear in our sample. We obtain the rest of the sample in the same fashion (with 000 corresponding to the farm numbered 1,000). Any number drawn that has already been obtained is ignored and the process continues until we have 100 different numbers (we do so in order to ensure that each farm is only included once in the survey).

This example is made easier by the fact that we have exactly 1,000 observations in the population. Suppose we had 1,350 farms. Using the last four digits of the second column, our first five observations would be 2,533, 4,805, 8,953, 2,529, and 9,970. None of these numbers appear in our list of farms! We could just discard these and continue until we had 100 different numbers in the range 1-1,350, but this would be both time consuming and tedious. A more practical way to proceed would be to imagine the four-digit numbers to be extracted as a combination of two two-digit numbers—note that of the first two digits, only the numbers 00, 01, 02, ..., 13 are of any use to us. Suppose we did the following for the first two digits of all numbers extracted:

Replace 00, 14, 28, ... , 84 by 00
Replace 01, 15, 29, ... , 85 by 01
. . .
. . .
Replace 13, 27, 41, ... , 97 by 13.

Numbers 98 and 99 are still ignored. The first four entries listed above would become 1,133, 0605, 0553, and 1,129 and we would then go to the next number (4,717, which would become 0517).

Alternatively, we could do the following:

Replace 1351, 2701, 4051, 5601, 6751, 8101 by 01
Replace 1352, 2702, 4052, 5402, 6752, 8102 by 02
.   .   .
Replace 2700, 4050, 5400, 6750, 8100, 9450 by 1350.

Numbers between 9451 and 9999 are still ignored. In this case, the first four entries would become 1183, 755, 853, 1179, while the last entry, 9970, is still discarded before going onto selecting a fifth entry.

Table 3 Some random numbers

10097	32533	76520	13586	34673	54876	80959	09117
37542	04805	64894	74296	24805	24037	20636	10402
08422	68953	19645	09303	23209	02560	15953	34764
99019	02529	09376	70715	28311	31165	88676	74397
12807	99970	80157	36147	64032	36653	98951	16877
66065	74717	34072	76850	36697	36170	65813	39885
31060	10805	45571	82406	35303	42614	86799	07439
85269	77602	02051	65692	68665	74818	73053	85247
63573	32135	05325	47048	90553	57548	28468	28709
73796	45753	03529	64778	35808	34282	60935	20344

APPENDIX 4

SELECTING CLUSTERS WHEN THEY ARE OF UNEQUAL SIZE

When clusters are of unequal size, the survey designer has three options: probability proportional to size (PPS), segmentation, and equal probability of selection method.

Probability Proportional to Size (PPS)

If the full sampling frame is at hand, or at least good approximations of the size of each cluster are available, then we may proceed by selecting the clusters with Probability Proportional to Size (PPS). An equal number of units can be then selected from each of the selected clusters in the second stage. This procedure will guarantee an equal probability of selection to all units, thus resulting in a self-weighting sample. The self-weighting feature of PPS sampling, combined with the possibility of drawing an equal number of elements from each cluster, makes the design advantageous from a practical standpoint. The following example illustrates the procedure.

Example: Imagine that our universe is composed by three villages of size 20, 30, and 50 households, respectively, and that we wish to draw a sample of 20 households. To cut on transportation costs and logistic difficulties, we may wish to limit the data collection to only two villages. To select the two villages using PPS, we order the villages and compute the cumulative size of the universe and the sampling intervals in which each numbered household falls into.

Village	Size	Cumulative size	Sampling Interval	Unit selected	Cluster selected
1	20	20	1-20		
2	30	50	21-50	37	X
3	50	100	51-100	88	X

(continued)

Example (continued)

We then randomly select two numbers between 1 and 100, the universe size. Random selection can be done by using tables of random numbers, or any other random selection technique. Suppose that the number 37 and 88 were drawn. From the last column, we can easily see that the two numbers selected fall into the 2nd and 3rd village. Thus, these two villages are selected. This method of selection is known as PPS, since the probability of a cluster being selected is proportional to its size (in our example, the probability that the first number drawn falls within the first village interval is 1/5, the second 3/10, and the third 1/2). Following this first-stage selection, we then proceed by randomly drawing 10 households in each of the two selected clusters. This design results in a self-weighting sample in which each household has an equal probability of selection of 1/100 (1/5 x 1/20 in the first cluster, 3/10 x 1/30 in the second, and 1/2 x 1/50 in the third, i.e., the probability of selecting the cluster times the probability to select a household within each cluster). Note that the equal probability of selection for all ultimate units derives from the fact that the difference in the sampling fractions between two clusters in the first stage of selection (biased in favor of larger clusters) is exactly compensated by the difference in sampling fractions between elements of the same two clusters in the second stage (biased in favor of elements in smaller clusters).

Segmentation

If the clusters are of variable but known size, a widely used alternative to PPS is to *segment* the clusters in equally-sized groups, and select with fixed probability a limited number of segments. Also, segmentation can be used to further reduce listing costs, by restricting the listing operations to only a portion of the original cluster. However, the segmentation procedure bears its own costs, which must be set off against the potential benefits. Sketch maps of the cluster, which include the location of the households as well as natural boundaries and easily visible landmarks, must be drafted before proceeding with the segmentation. It is often suggested that the size of the segments are determined on the basis of the chosen take. In the above example, we could have divided each cluster into segments of size 10, then select two segments for complete enumeration.

Segmentation becomes more complicated as segments get smaller, since they will be less likely to match natural boundaries, if based on area mapping. In addition, the smaller the area comprising the segment, the higher the sampling error, since adjacent units are more likely to exhibit characteristics that are correlated.

Equal Probability of Selection Method (EPSEM)

If a frame is not obtainable, or if information on the size of the clusters is not available or not considered reliable, the use of a multistage equal probability of selection method (EPSEM) may be more feasible than the two methods presented above. The procedure consists of randomly selecting a restricted number of clusters irrespective of their size. A limited enumeration of the selected groups must take place to allow the draw of a number of ultimate units in each cluster proportional to its size.

> **Example:** Using the previous example, assume two villages are drawn using SRS. If the selected villages are the 2nd and the 3rd, then, using SRS or systematic sampling, we would draw 8 (= 30/80 x 20) and 13 (= 50/80 x 20) units from each cluster, respectively, i.e., a number proportional to the size of the cluster.

The main advantage of this method over PPS when the cluster size is unknown is that it eliminates the need for weighting the observations in the different clusters, while still maintaining the cost-saving and limited enumeration of the two-stage design. The main drawback of EPSEM is that the number of ultimate units selected out of each cluster is different. In most circumstances, it is likely to represent a considerable logistic burden. In addition, it may not be very cost-effective, particularly if some of the cluster takes—the number of households or units of observation within the cluster—are particularly small.

A less viable alternative to any of these multistage designs would have been to select the 20 households directly from the whole universe using SRS. The obvious drawback of this method is that it is likely that the 20 households will belong to all three village, increasing the cost of the survey. In addition, very few households may happen to be drawn from one or two villages, making the increase in transportation costs not worth the gain in efficiency associated with SRS. The potential improvement in efficiency associated with the single-stage design may actually have to come at the cost of a reduction in the sample size. In practice, because transportation costs have gone up, a lower number of households may be interviewed. The existence of this trade-off may ultimately result in a less efficient estimate compared to the two-stage design.

REFERENCES

Bernard, H. R. 1988. *Research methods in cultural anthropology*. Newbury Park, Calif., U.S.A.: SAGE Publications.

Casley, D. J., and K. Kumar. 1988. *The collection, analysis, and use of monitoring and evaluation data*. Baltimore, Md., U.S.A.: Johns Hopkins University Press.

Casley, D. J., and D. A. Lury. 1987. *Data collection in developing countries*. Oxford: Clarendon Press.

Cochran, W. G. 1997. *Sampling techniques*. New York: John Wiley.

Devereux, S., and J. Hoddinott, eds. 1992. *Fieldwork in developing countries*. London and Boulder, Colo., U.S.A.: Harvester Wheatsheaf and Lynne Rienner.

Kish, L. 1965. *Survey sampling*. New York: John Wiley.

Newbold, P. 1988. *Statistics for business and economics*, 2nd edition. Englewood Cliffs, N.J., U.S.A.: Prentice Hall.