Assessing the Relative Poverty of Microfinance Clients

A CGAP Operational Tool

CARLA HENRY
MANOHAR SHARMA
CECILE LAPENU
MANFRED ZELLER

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
2033 K STREET, N.W.
WASHINGTON, D.C. 20006, USA
TEL. (202) 862-5600
FAX. (202) 467-4439

JULY 2000

CONTENTS

PART III: Applying the Results

List of Annexes

List of Figures

List of Tables

ACKNOWLEDGMENTS

This handbook is the result of a year-long collaboration between the Consultative Group to Assist the Poorest (CGAP) and The International Food Policy Research Institute (IFPRI) to develop a standardized tool to measure the poverty of microfinance institution (MFI) clients. The project was financed by CGAP and the methodology for the tool developed by IFPRI.

The authors want to express their gratitude to CGAP for its commitment to improving the performance of microfinance institutions through development of better tools to appraise MFI performance. It is from this commitment that the idea of a poverty-assessment tool was borne and funding was made available to develop such a tool. Thanks is also due to the four case-country MFIs for their cooperation in testing the tool by sharing information and opening access to their clients: in India, SHARE; in Kenya, the Kenya Women Finance Trust; in Nicaragua, ACODEP; and in Madagascar, Desjardins.

IFPRI's collaborating researchers in each of the four countries played a key role in testing and refining methodology: in Madagascar, the National Center for Research on Rural Development (FOFIFA); in Nicaragua, independent consultant Ms. Delia Maria Sevilla Boza; in India, the National Institute of Rural Development; and in Kenya, the Tegemeo Institute of Egerton University. Thanks to the high caliber of their work, we rest assured that local research institutes will manage well and benefit from opportunities to apply this tool for assessing the poverty outreach of other MFIs.

The authors also thank the following persons who reviewed an earlier draft and provided detailed and insightful comments: Edwardo Bazoberry, Prodem, Bolivia; Monique Cohen, USAID; James Copestake, Univ. of Bath; Alex Counts, Grameen Foundation; John deWit, Small Enterprise Foundation, South Africa; Gary Gaile, University of Colorado, and David Hulme, Manchester University. In IFPRI, Marinella Yadao and Jay Willis provided significant assistance in preparing the document.

Finally, the authors would like to record special thanks to Brigit Helms and Syed Hashemi, CGAP, who were involved with this project right from the start, and who provided critical inputs at all stages of the study.

FOREWORD

This manual describes an operational tool for assessing the poverty of clients of microfinance institutions (MFIs). The methodology outlined is intended for use by MFI evaluators as a practical, accurate, and relatively simple means of assessing the extent to which MFI programs reach the poor.

The Consultative Group to Assist the Poorest (CGAP) commissioned the manual's development in recognition of donor needs for improving the transparency of MFIs in the degree to which they reach the poor. The manual complements an array of tools already available for assessing the institutional performance of MFIs. CGAP's companion manual, *Format for Appraisal of Microfinance Institutions*, outlines a framework for assessing the likelihood that the MFI will develop into an entity capable of expanding independent of donor resources. However, it is widely perceived that a trade-off exists between financial sustainability and depth of outreach to the poor. The combination of these two assessment tools can provide a straightforward means of gauging the likelihood of an MFI's reaching poor clients while relying on predominantly commercial funding.

The poverty-assessment tool is relatively easy to implement in a short time and at minimum cost to the donor organization, both of which were set as criteria for the tool's development. In addition, the tool supports comparison of poverty outreach among MFIs and even across countries. The methodology is applicable to all MFIs, regardless of their location, client structure, and outreach approach.

The methodology was field tested through four case studies in Asia, Africa, and Latin America conducted during 1999. The cumulative experience gained from these studies provided insight into how the tool could be standardized while still maintaining adaptability to local conditions.

The Authors

CHAPTER ONE:
POVERTY ASSESSMENT OF MICROFINANCE INSTITUTIONS

The microfinance industry promotes the dual objectives of sustainability of services and outreach to the very poor. When deciding to fund specific microfinance institutions (MFIs), donors and other social investors in the sector invest in both objectives, however their relative importance varies among funders. Furthermore, many practitioners, donors, and experts perceive a trade-off between financial sustainability and depth of outreach, although the exact nature of this trade-off is not well understood.

In recent years, several tools have emerged to assist donors in their assessment of the institutional performance of MFIs. An example is the CGAP *Appraisal Format*. This latter tool contains practical guidelines and indicators for measuring MFI performance in a range of issues, including: governance, management and leadership, mission and plans, systems, operations, human resource management, products, portfolio quality, and financial analysis. Analysis of these institutional features allows for an appraisal of the potential for institutional viability or sustainability. At the same time, the proliferation of tools such as the *Appraisal Format* has encouraged transparency and the development of standards on the topic of financial sustainability.

Currently, no concrete tool for measuring the poverty level of MFI clients exists. In order to gain more transparency on the depth of poverty outreach, CGAP collaborated with the International Food Policy Research Institute (IFPRI) to design and test a simple, low-cost operational tool to measure the poverty level of MFI clients relative to nonclients. This tool comprises a companion piece to the CGAP *Appraisal Format* and donors should not use it in isolation from a larger institutional appraisal.

IFPRI developed a survey-based method of assessment and tested it with case studies using random samples of client and nonclient households from the operational areas of four CGAP partner MFIs. Not only did these institutions operate in significantly different geographic and socioeconomic settings, they also differed in terms of their objectives and institutional design. A sample of 500 households—200 client households and 300 nonclient households—were drawn in each of the case studies. Results from these case studies helped refine the final product, a practical operational manual. This manual explains in detail the process for conducting a comparative poverty assessment between MFI clients and nonclients.

1.1 Study parameters and choice of an indicator-based methodology

1.1.1 The scope of the assessment tool

The immediate objective of the research project directly influenced the assessment method adopted: to develop a tool that could be used by CGAP and other donors to assess the poverty level of microfinance clients. In order for the tool to be effective and practical, the tool needed to have the following features:

- The methodology should be simple enough to remain operational;
- The methodology used should permit comparison between different MFIs and, if possible, across countries; and
- The tool should not be costly to implement and should have a minimum turnaround time without sacrificing too much in terms of credibility of results.

Consideration of these parameters led to the adoption of the indicator-based method. This method involved the following main tasks:

1. Identifying a range of indicators that reflect powerfully on poverty levels, and for which credible information can be quickly and inexpensively obtained;
2. Designing a survey methodology that facilitates the collection of information on these indicators from households living in the operational area of the MFI; and
3. Formulating a single summary index that combines information from the range of indicators and facilitates poverty comparisons between client and nonclient households.

Approaches based on intensive households expenditure surveys were ruled out not only because they were too expensive and time-consuming to implement, but also because they necessitated advanced skills in statistical data analysis. On the other hand, participatory or rapid assessment techniques were ruled out mainly because they did not easily allow for objective comparisons between MFIs. A brief discussion of these alternative approaches is given in Annex 1.

1.1.2 Methodological steps using the indicator-based approach

The indicator-based approach involved the following methodological steps:

1. Extensive literature review and expert consultation on the general availability and use of poverty indicators,
2. Selection of indicators based on an eight-point criteria,
3. Development of a generic questionnaire for testing in the four case studies,
4. Adaptation of the questionnaire to account for local-level specificities using participatory methods,
5. Testing indicators through household surveys,
6. Statistical analysis of indicators,
7. Review of indicators with MFI and other stakeholders,
8. Selection and synthesis of common indicators across countries,
9. Development of a generic poverty index,
10. Revision and simplification of generic questionnaire.

1.1.3 Multiple dimensions of poverty and its implication

Because of the multifaceted nature of poverty, reliance on any one dimension or any one type of indicator was not recommended. To capture different dimensions of poverty, IFPRI used the following general classification of indicators in the process of developing the generic questionnaire:

1. Indicators expressing the means to achieve welfare. These reflect the earning potential of households and relate to:

- Human capital (family size, education, occupation, etc.),
- Asset ownership,
- Social capital of household.

2. Indicators related to the fulfillment of basic needs:

    - Health status and access to health services,
    - Access to food, shelter and clothing.

3. Indicators related to other aspects of welfare (security, social status, environment).

In many cases, a single indicator may not be fully reliable even to describe one particular dimension of poverty. For example, collecting information on ownership of a TV is not likely to shed complete light on a household's access to consumer assets in general, and needs to be supplemented by other indicators on ownership of kitchen appliances and/or other electronic assets such as radios or electric fans.

1.1.4 Criteria for selection of indicators

From an exhaustive list of indicators obtained through a literature review, the IFPRI team initially chose to include a smaller subset in the generic questionnaire. The criteria used in their selection include:

- Nationally valid (can be used in different local contexts, urban vs. rural),
- Not too sensitive a question (can be asked openly),
- Practical (can be observed as well as asked),
- Quality of the indicator (discriminates poor households individually),
- Reliability (low risk of falsification/error; also possible to verify),
- Simplicity (direct and easy to answer vs. computed information),
- Universality (can be used in different countries).

A list of indicators with their rankings based on this criteria is provided in Annex 2.

1.2 Development of a generic questionnaire

Based on extensive analysis of the initial long list, IFPRI included the following types of indicators in the generic questionnaire to test in the four case studies:

- Demographic characteristics of household and members (e.g., family size, age and number of children),
- Quality of housing (eg. walls, roofs, access to water),
- Wealth (eg. type, number and value of assets),
- Human capital (eg. level of school education and occupation of household members),
- Food security and vulnerability (eg. hunger episodes in last 30 days/12 months, types of food eaten in last two days),
- Household expenditures for clothing (poverty benchmark).

1.2.1 Purpose of field-testing

The questionnaire was field tested in each of the four case studies with the following objectives in mind:

1. To further select and/or reduce the number of indicators to include in the recommended final questionnaire by taking the following steps:

    - In each case study, identify indicators that are tightly related to poverty levels;
    - Identify indicators that can be commonly used across the four countries (that is, those that are robust to diverse socioeconomic and cultural contexts);
    - Identify indicators suitable for capturing local specifities and evaluate their importance in overall assessment;
    - Catalogue problems and strengths of the survey tool and related analysis through testing in different country and MFI settings; and,
    - Share results with MFIs and other stakeholders to critically evaluate the method.

2. To test and standardize the method to integrate different indicators into a poverty index that allows comparisons between MFIs and countries.

3. To document all procedures involved in (1) and (2) in a user-friendly manual to support future independent assessments.

1.2.2 Indicators in the final recommended questionnaire

Table 1.1 lists indicators included in the final recommended questionnaire. (A copy of the final recommended questionnaire is included as Annex 3.) Their selection was based on 1) the ease and accuracy with which information on them could be elicited in a typical household survey, and 2) how well they correlated with the benchmark poverty indicator: per capita expenditure on clothing and footwear. Per capita expenditure on clothing and footwear was chosen as the benchmark indicator since it bears a stable and highly linear relationship with total consumption expenditure, a comprehensive measure of welfare at the household level.

The following indicators were rejected:

- Indicators using child-specific information. Not all households have children; hence using child-related information precluded some households from comparative analysis.
- Indicators of social capital. This is an evolving area of investigation, and measurable and comparable indicators were not easily found.
- Subjective responses. Responses on self-assessment of poverty were considered unreliable to be used in comparisons
- Health related information. Eliciting health-related information requires longer recall periods and more intensive and specialized training of interviewers. In the absence of training

provided by health specialists (which is expensive), responses can be highly subjective and misleading.

TABLE 1.1 Indicators in the final recommended questionnaire

Human Resources	Dwelling	Food security and vulnerability	Assets	Others
• Age and sex of adult household members • Level of education of adult household members • Occupation of adult of members of household • Number of children below 15 years of age in the household • Annual Clothing/foot-wear expenditure for all household members	• Ownership status • Number of rooms • Type of roofing material • Type of exterior walls • Type of flooring • Observed structural condition of dwelling • Type of electric connection • Type of cooking fuel used • Source of drinking water • Type of latrine	• Number of meals served in the last two days • Serving frequency (weekly) of three luxury foods • Serving frequency (weekly) of one inferior food • Hunger episodes in last one month • Hunger episodes in last 12 months • Frequency of purchase of staple goods • Size of stock of local staple in dwelling • Marginal propensity to consume out of additional income	• Area and value of land owned • Number and value of selected livestock resources • Ownership and value of transportation-related assets • Ownership and value of electric appliances	• Urban/rural indicator • Nonclient's assessment of poverty outreach of MFI

1.3 Methodology overview

1.3.1 Using principle component analysis to develop the poverty index

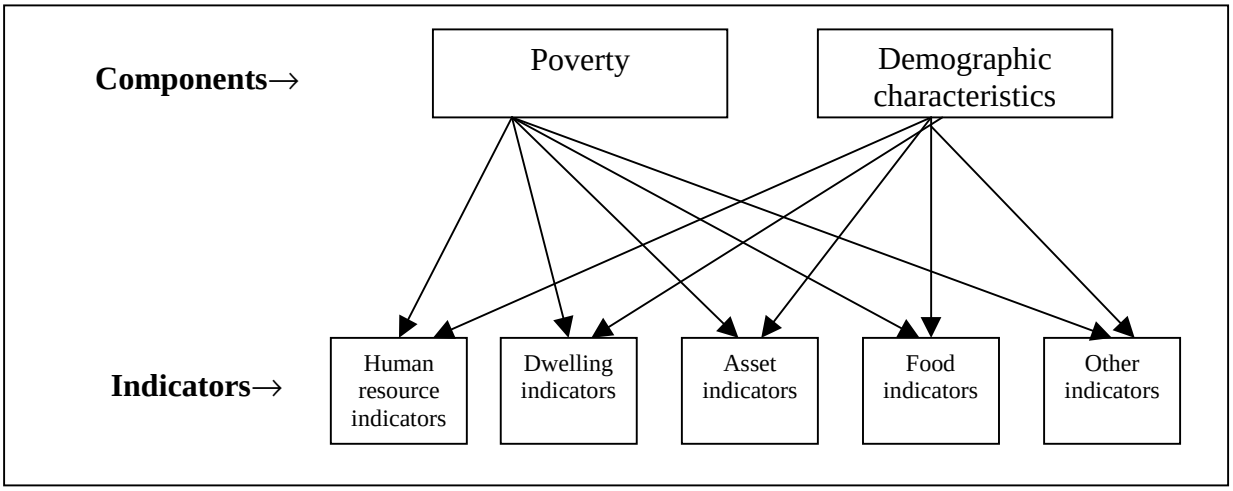
The use of multiple indicators enables a more complete description of poverty, but it also complicates the task of drawing comparisons. The wide array of indicators has to be summarized in a logical way, underlining the importance of combining information from the different indicators into a single index. The creation of an index requires finding a method of weighting that can be meaningfully applied to different indicators so as to come to an overall conclusion.

The case studies used the method of principle component (PC) analysis to accomplish this task. Specifically, PC analysis isolates and measures the poverty component embedded in the various poverty indicators and creates a household-specific poverty score or index. Relative poverty comparisons are then made between client and nonclient households based on this index.

PC analysis extracts underlying components from a set of information provided by indicators. In the case of this assessment tool, information collected from the questionnaires make up the "indicators" and the underlying component that is isolated and measured is "poverty."[1]

In the example presented in Figure 1.1, poverty and demographic characteristics constitute the two underlying components affecting the level of all the indicators. Because the indicators are determined by these common underlying components, they are likely to be related to each other. PC analysis uses this information (the co-movement amongst the indicators) to isolate and quantify the underlying common components. PC analysis is also used to compute a series of weights that mark each indicator's relative contribution to the overall poverty component. Using these weights, a household specific poverty index (or poverty score) can be computed based on each household's indicator values.

FIGURE 1.1 Indicators and underlying components



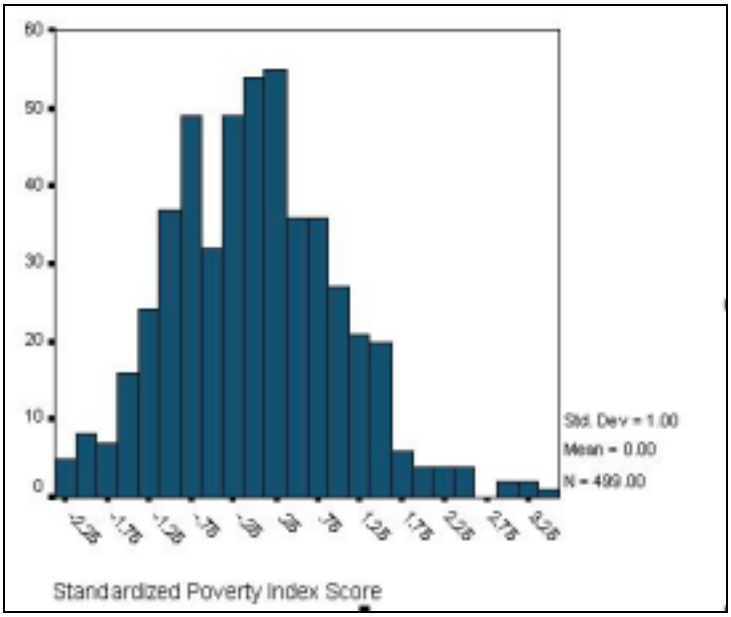
The indicators in the case studies were specially chosen to correlate well with poverty, including only those that had significant correlation with per capita clothing expenditure, the benchmark indicator. Hence the poverty component is expected to account for most of movements in the indicators, and will be the "strongest" of all the components. Further, the poverty component is also identified based on the size and consistent signs of the indicators in their contribution to the index. For example, education level should contribute positively – not negatively – to wealth.

[1] The principal component technique slices information contained in the set of indicators into several components that have the following characteristics:

1. Each component is constructed as a unique index based on the values of all the indicators. This index has a zero mean and standard deviation equal to one.
2. The first principal component accounts for the largest proportion of the total variability in the set of indicators used. The second component accounts for the next largest amount of variability not accounted by the first component, and so on for the higher order components. In our case, therefore, the first principal component will be the poverty component.
3. Each component is unrelated to the other components; that is, each represents a unique underlying attribute.

The principle component analysis produces a household-level poverty index. Figure 1.2 gives an example of the distribution of the poverty index across households using MFI B data. The greater the value of the score, the relatively wealthier the household.

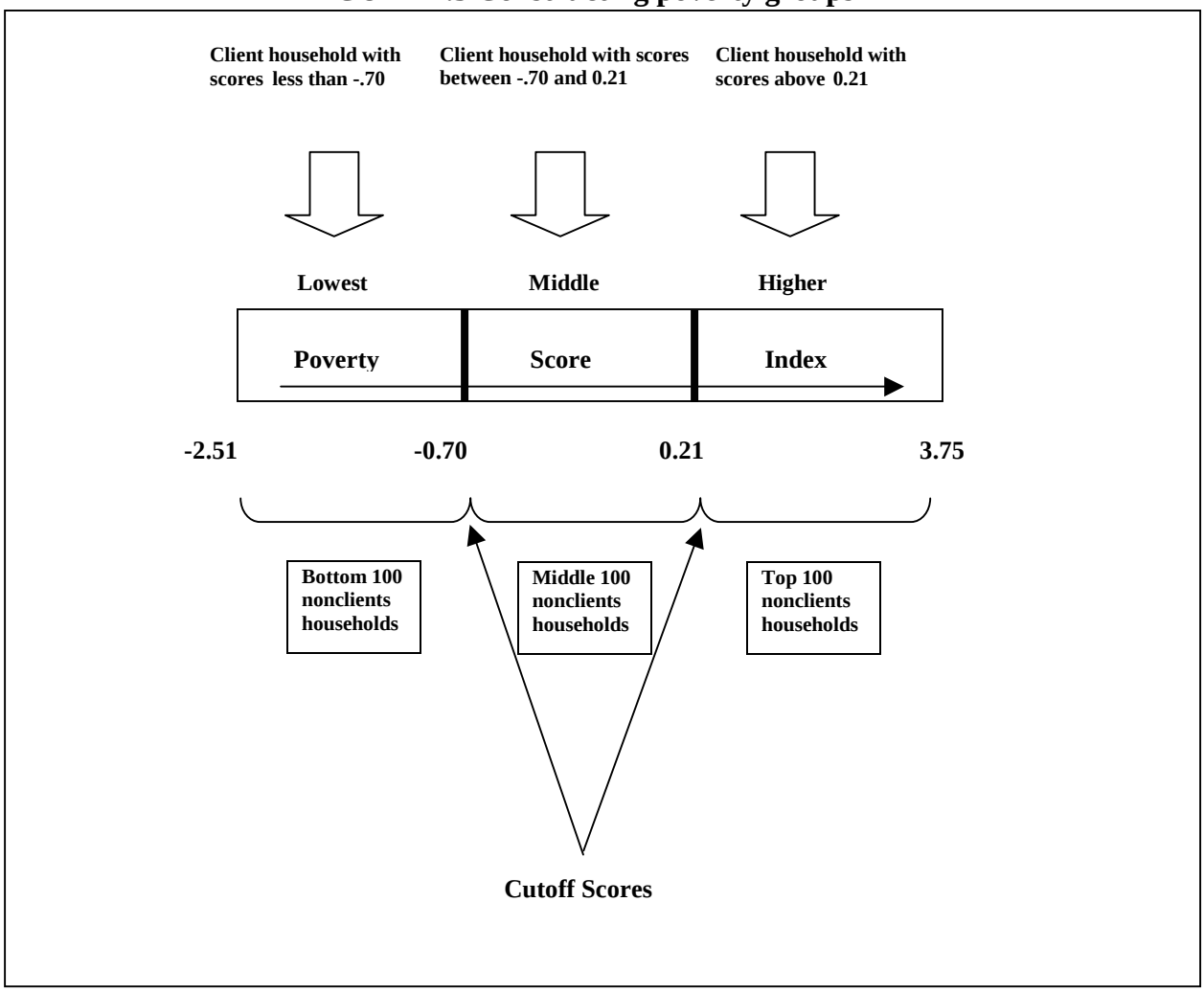
FIGURE 1.2 Histogram of the standardized poverty index (MFI B)



1.3.2 Using the poverty index

Each assessment study includes a random sample of 300 nonclient households and 200 client households. To use the poverty index for making comparisons, the nonclient sample is first sorted in an ascending order according to its index score. Once sorted, nonclient households are divided in terciles based on their poverty index score: the top third of the nonclient households are grouped in the "higher" ranked group, followed by the "middle" ranked group and finally the bottom third in the "lowest" ranked group. Since there are 300 nonclients each group contains 100 households each. The cutoff scores for each tercile define the limits of each poverty group. Client households are then categorized into the three groups based on their household scores. Figure 1.3 illustrates the use of cutoff scores to create poverty terciles from nonclient households. The cutoff scores of -.70 and +.21 were calculated from the case study example shown in Figure 1.2.

FIGURE 1.3 Constructing poverty groups



If the pattern of client households' poverty matches that of the nonclient households, client households would divide equally among the three poverty groupings just as the nonclient households, with 33 percent falling in each group. Hence any deviation from this equal proportion signals a difference between the client and the nonclient population. For instance, if 60 percent of the client households fall into the first tercile or lowest poverty category, the MFI reaches a disproportionate number of very poor clients relative to the general population.

1.3.3. Relative vs. absolute poverty

The poverty index provides a tool to calibrate *relative* poverty—the extent to which a household is *worse off* or *better off* compared to other households. It does not by itself provide information on the *absolute* level of poverty, the actual level of depravation of the "lowest" category of households or the level of affluence of the "higher" group. A good sense of absolute level of poverty among clients and nonclients can be got by noting and comparing values of individual indicators themselves. Chapter 7 describes how to do this. Another assessment of the absolute poverty can be got from

comparing welfare indicators at the national level, indicators such as per capita real incomes or the human development index (HDI). How results from the analysis of the poverty index should be juxtaposed with regional and national level indicators in making final inferences is illustrated in section 1.4 and described in detail in Chapter 9.

1.4 Interpreting results

1.4.1 Selected case study MFI results

The results are best summarized by examining the proportion of client households falling into the three poverty groups. If the pattern of client households' poverty were similar to those of the nonclient households, client households would divide up equally among the three poverty groupings. Any deviation from this proportion signals a difference between the client and the nonclient population. Below is a summary of results from the four case studies used to test the methodology.

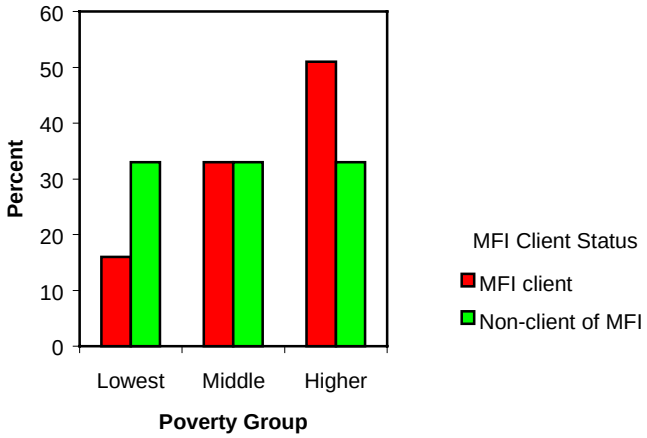
MFI A. Figure 1.4 presents the poverty groups by client and nonclient households. The distribution of MFI A's clients across the poverty groups closely mirrors the distribution of nonclients, indicating that MFI A serves a clientele that is quite similar to the general population in its operational area. This result is consistent with MFI A's stated objective of reaching micro, small, and medium enterprises and the diversity in the financial products that it offers.

FIGURE 1.4 MFI A: Distribution of client and nonclient households across poverty groups

Relative poverty group	% Client households	% Nonclient households
Lowest	31	33
Middle	38	33
Higher	31	33

MFI B. Figure 1.5 shows that the poorest households are underrepresented among MFI B clients. However, about one-half of the clients fall into the two poorest categories, which is remarkable considering the mission of the institution (to reach all women in business), the focus of the product (to finance businesses after submitting a business plan), and the lack of overt targeting.

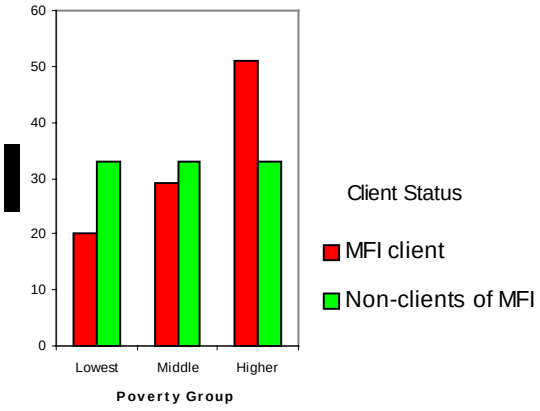
FIGURE 1.5 MFI B: Distribution of client and nonclient households across poverty groups



Relative poverty group	% Client households	% Nonclient households
Lowest	16	33
Middle	33	33
Higher	51	33

MFI C. About half of MFI C's clients belong to the 'upper' poverty group while they are underrepresented in the lowest poverty group (Figure 1.6). This result reflects the fact that MFI C's membership is share-based and open to all individuals. However, poverty outreach is significantly higher when considering only clients belonging to the new program for women. Nearly one-half (45.2 percent) of these clients belonged to the 'poorest' group, and only 19 percent of the new women clients belong to the 'less poor' group.

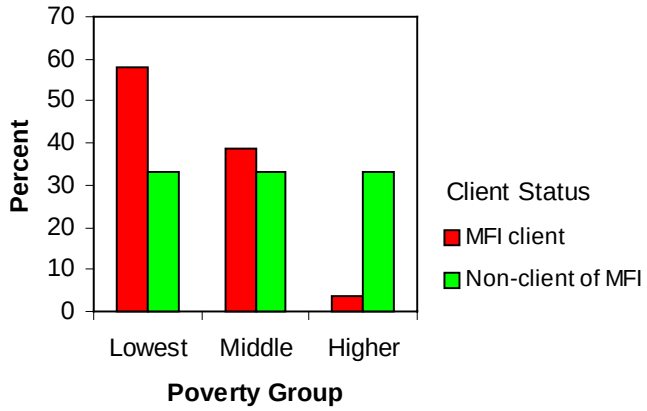
FIGURE 1.6 MFI C: Distribution of client and nonclient households across poverty groups



Relative poverty group	% Client households		% Nonclient households
	Typical clients	Women's program	
Lowest	20	45	33
Middle	29	36	33
Higher	51	19	33

MFI D. Figure 1.7 indicates quite clearly that the poorest groups are strongly overrepresented and that less poor households are underrepresented among MFI D's clients. This result is not only consistent with MFI D's explicit aim to serve the poorest households in its operational area but also indicates considerable success in its targeting practices.

FIGURE 1.7 MFI D: Distribution of client and nonclient households across poverty groups



Relative poverty group	% Client households	% Nonclient households
Lowest	58	33
Middle	38.5	33
Higher	3.5	33

1.4.2 Overall comparative results

A comprehensive assessment of an MFI must include an evaluation of how its poverty outreach record reconciles with its mission and program objectives. As the case studies themselves have shown, MFIs differ in terms of geography, their stated mission, the type of market niche they seek, their preference for a specific type of institutional culture, and a host of other factors. Ignoring these considerations or providing incomplete information on institutional details fails to tell a complete story and the method can be easily misused. With this important caveat, a basis for making overall comparisons across MFIs and countries is discussed below.

Table 1.2 presents three ratios that facilitate comparisons between MFIs. Ratio 1 is computed by dividing the percentage of client households that belong to the poorest group by 33, the percentage of nonclient households that belong to this group. The ratio reflects the extent to which the poorest households are represented in the client population.

A ratio of one indicates that the proportion of the poorest households among the MFI's client equals that of the general population. Ratios higher than one imply that the proportion of the poorest households among the MFI's clients exceeds that in the general population. On the other hand, ratios less than one imply that the proportion of the poorest households among the MFI's clients falls below that of the general population.

A similar ratio—Ratio 2—divides the percentage of client households that belong to the higher poverty group by 33. The ratio reflects the extent to which less poor households are represented in the client population. A ratio above one indicates that, in comparison to the nonclient population, a greater proportion of client households fall into the 'higher' ranked poverty group.

While Ratios 1 and 2 provide relative poverty comparisons in the operational area of the MFI, this information must be supplemented by country-level information using the human development index (HDI) computed by UNDP. Two of the four case study countries fall below the all-developing country average, and the human development index for the Southern African country where MFI C is located equals less than 11 percent of the average for all developing countries taken together. Therefore, even the 'higher' ranked clients of MFI C are likely to be very poor according to international standards.

TABLE 1.2 Relative poverty ranking of client vs. nonclients

Percentage/ Ratio	MFC A	MFC B	MFC C	MFC D
Percent of client households who are as poor as the poorest one-third of the nonclient population	30.9%	20.3%	16%	58%
Ratio 1	0.94	0.62	0.48	1.76
Percent of client households who are as well of as the least poor one-third of the nonclient population	31.4%	50.8%	51%	3.5%
Ratio 2	.95	1.54	1.55	0.11
Ratio of country HDI to HDI for all developing countries taken together	1.16	0.94	0.89	1.04

1.5 Intended users

Donors are the intended beneficiaries of this manual, although they are not envisioned as the actual implementers of the tool. The manual presents as simply as possible the techniques involved in conducting the poverty assessment, but the implementation process remains best handled by a team of research professionals with expertise in survey methodology and statistical analysis. In almost all countries, knowledgeable and reliable research institutes regularly conduct studies that involve levels of detail similar to that presented in this manual. By documenting all steps of the survey design, data collection, and analysis, as well as the interpretation and reporting of results, this manual is a clear-cut guide for the experienced researcher to conduct the poverty assessment.

Donors will want to read through the manual to develop an understanding of the level of effort and timeframe required, the likely costs associated with the assessment, and the level of expertise to look for within a contracting institute. The manual provides specific guidelines for contracting individuals or institutions to conduct the assessment. The assessment is intended to be conducted independent of the MFI's direct involvement, but the manual will also indicate the types of information support required from the MFI. Donors will also want to review the results of using the tool to anticipate how the quantitative measurement of poverty outreach can best be integrated into additional appraisal methods.

The tool is also not meant for direct use by the MFI. This is not only because the level of specialized knowledge is unlikely to be found among MFI staff, but also because direct field testing by the MFI could greatly bias household responses. The results of the assessment tool will likely interest MFIs, and MFIs may even have ideas on how the results can be used for their own purposes. However, the assessment tool is not intended to guide MFIs in applying assessment results to their future program

development. Any decision on how to make use of the assessment results is left solely to the MFI and donor.

In many cases, MFI's and donors will find other relative poverty tools fit better their operational activities. These alternative tools serve ongoing needs in monitoring the poverty levels of new clients, or operational tools to screen new applicants. Other tools compare the poverty status of clients more directly to a country's estimated poverty line.

1.6 Manual layout

The manual is divided into three parts. Part I, Collecting survey Data (Chapters 2-4), provides guidelines and instruction for collecting survey data. Chapter 2 guides users in developing a sampling frame and in actually sampling of households. Chapter 3 outlines how to customize a standardized questionnaire to fit the specific local conditions where the MFI operates. Chapter 4 describes the level of detail needed to successfully implement the survey and gives guidelines for contracting the project to a qualified institution or individual.

Part II, Analyzing Data (Chapters 5-8), focuses on managing and analyzing the data using the Statistical Program for Social Science (SPSS) software. Chapter 5 guides users in managing the survey data once collected, including how to enter data, the structure of files and how to link files, and how to clean data. Chapter 6 summarizes SPSS techniques for working with the data to prepare it for analysis. Chapter 7 gives an overview of data analysis techniques used to describe socioeconomic similarities and differences between the survey households and how to use SPSS to implement them. Chapter 8 gives users an overview of the statistical procedure and the principle component analysis used to create the poverty index, and describes each step in creating a poverty index through application of principle component analysis technique.

Part III, Interpreting Results (Chapter 9), explains how results of the analysis can be applied by donors to form conclusions.

Figure 1.8 outlines the stages described in this manual to develop a poverty index and apply it to assess the relative poverty of MFI clients.

FIGURE 1.8 Stages to implement the poverty assessment tool

Stage one:	Sample through random methods that support results being interpreted as representative of the MFI client and nonclient populations.
Stage two:	Develop a formalized questionnaire that is adapted from a standardized template to fit local conditions.
Stage three:	Minimize risk of error through well-prepared survey implementation.
Stage four:	Apply standardized techniques for managing and analyzing data that ensure consistent and appropriate use and interpretation of data.
Stage five:	Interpret quantitative results using standardized measures to compare across countries and programs.

Stage six: Evaluate results within a more qualitative review of the MFI.

PART I

COLLECTING SURVEY DATA

CHAPTER TWO:
DEVELOPING THE SAMPLE DESIGN

The poverty-assessment tool is based on a comparison of relative poverty levels of MFI new-client households with nonclient households. The sampling design process determines how to select randomly a representative number of client and nonclient households. Through random sampling, the results found for the sampled households can be seen to hold true for the entire population of MFI client and nonclient households living in the MFI's operational area.

The researcher follows a series of steps for ensuring that the final set of households surveyed represents a random sample of all possible households that could have been interviewed. These steps are described in detail below.

2.1 Step one: Define the population and sampling unit

2.1.1 The household as the basic sampling unit

While some characteristics of poverty can be measured at the individual level, such as a person's income or assets only he or she owns, much of an individual's wealth is shared with and influenced by the household in which that individual lives. Assessing the relative poverty of an individual without considering the conditions of the entire household provides a distorted view. Therefore, this tool measures the relative poverty of the household rather than that of any single member in the household.

The approach has the disadvantage of not accounting for uneven distribution of household wealth within the household. MFI's targeting disadvantaged household members may have a stronger poverty outreach than indicated if the targeted members have limited access and control over a disproportionately small share of household resources.

Client households

Consider only clients newly joining the MFI. Only households of new MFI clients are considered eligible for this survey on the basis of the assumption that their living standards have not yet been affected by MFI participation. Every attempt should be made to capture as new a sample of clients as possible. A pool of clients in queue to receive the first round of services would be the ideal group to draw the sample from. When this is not possible, a general rule to follow is to define a new client as someone who has joined the MFI within the past six months, and has not yet received a loan from the MFI being assessed. If the MFI has a large number of new clients, the rule can be tightened to sample only clients joining in the previous three months.

Adopt filtering criteria that respond to the situation at hand. In specific cases, new-client selection criteria may prove too stringent—MFIs may not know how long a client has been with the MFI, or which new clients have already received loans. In addition, an MFI may accept new clients

only on a yearly basis, as is the case for many agriculture-based lending schemes. In general, if the sampling process is not able to rule out new clients who have already received a loan, additional precautions are called for in the questionnaire to control for possible effects from receiving this loan. These are discussed in the next chapter.

If necessary, eliminate new clients joining older groups. A further restriction on the type of new-client households eligible for sampling may be to exclude individuals who have recently joined client groups in existence longer than six months. This filter is often necessary because information about the number of new individuals in older groups is unavailable at a central location, or may be unreliable. In addition, surveying new members in older groups, who are likely to be few in number but spread over large areas, may prove logistically too costly.

Check that all household members meet the criteria. Client households may have more than one member currently active in the MFI as long as none have been clients for longer than six months. In addition, the household should not have any member who was once an MFI client but is no longer active.

Nonclient households

The sampling of nonclient households also requires that no household members be current or past clients of the MFI being assessed. Both clients and nonclients can be active participants in other MFIs and still be eligible for the sample.

2.1.2 Determining a feasible population area

Determine the operational area of the MFI. In addition to knowing which households to sample, the sampling area must also be determined. The operational area is the geographical area in which the MFI operates.

Identify any problem areas that may not be feasible to survey. In general, the research team will need to determine a standardized rule or set of rules for filtering out unfeasible survey areas and then follow this rule consistently. However, the process for eliminating areas from consideration needs to be carefully scrutinized to avoid any unintentional introduction of bias. Bias could occur if the remote areas screened out actually constituted relatively poorer households than those found in sampled areas. In addition, if the MFI operates different outreach programs, or offers a different basket of services to more remote regions, these differences would not be captured.

Before setting any rules to limit the feasible area for the survey, the MFI needs to be asked about possible selection bias associated with the research team's proposed rules. In addition, the MFI may also have reasons for excluding operational areas from the survey. In some cases, these reasons need to be respected, such as if the survey is likely to raise local animosity to the MFI. In other cases, these reasons may cover up other motives that the MFI does not wish to make clear. In summary, determining the feasible area for the survey requires good information and careful judgment to avoid bias.

Document any potential source of bias from limiting the feasible area. Exclusion of infeasible areas may introduce a bias if the areas excluded are likely to be either below or above the poverty levels found in remaining areas. In a recent assessment, densely populated urban areas were excluded because households here would not have been willing to participate in the survey. Their exclusion was noted and factored in when assessing the regional coverage of the MFI. Exclusion of some areas should not introduce bias of local results since MFI clients and nonclients in these areas are equally excluded.

2.2 Step two: Construct the MFI-based sampling frame

The sampling frame refers to how information on the number and distribution of new MFI clients within the MFI's operational area can be used to determine which localities will be surveyed. The sampling frame is a list or a plan for accounting for all new clients and nonclients located in the areas where the MFI currently accepts clients. If the MFI's operational area has been limited for budget or logistical reasons, the sample frame would cover only this reduced area.

In most cases, the research team need not compile an actual list of all qualifying MFI new client households. Instead, a set of procedures can be followed to randomly select survey localities. Only in these selected survey localities will client lists eventually be formed. Determining the locations of these survey localities will require the use of several sampling techniques, as described below.

2.2.1 Cluster sampling for MFI new clients

Cluster sampling is a sampling technique that randomly reduces the number of MFI localities to be eventually surveyed. Area clusters are formed based on the geographical divisions the MFI has already created. These are usually subdivisions of MFI operational regions or branches. In many cases, the sub-areas defined for each MFI field agent can be treated as a separate cluster if these do not overlap. Cluster sampling requires that the entire feasible area for the survey be divided into non-overlapping clusters with the intent that only a subset of these clusters will be randomly chosen for the actual sampling of households. Deciding how to form clusters will largely be determined by the MFI geographical delineations defined to establish sub-areas of operation. As a rule, the more clusters that can be identified, the better. A minimum of 10 clusters is recommended, but a number closer to 30 is preferred. From these clusters, it is recommended to randomly select only five to six clusters for actual sampling of households.

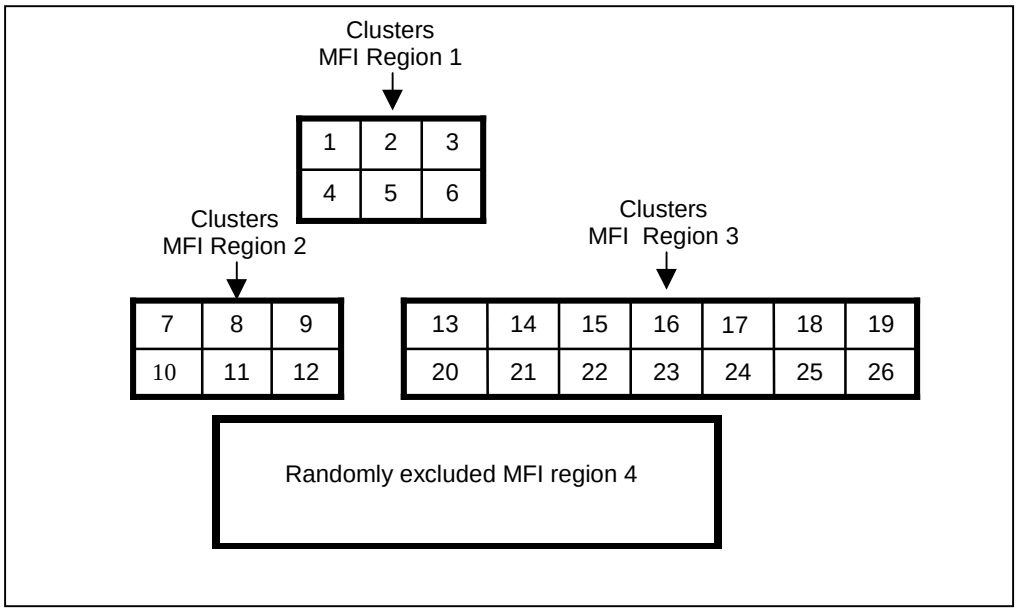
Without clustering MFI clients, simple random sampling could result in a list of survey households scattered over too large of an area to manage either financially or logistically. In general, clustering is required for MFIs that

- have large operational areas,
- have very large numbers of new clients, or
- are geographically very scattered so that travel costs would be high.

Because most MFIs fill at least one of these conditions, we recommend the cluster approach as the default method for reducing the number of sampling areas in a sample frame.

Finally, for very large MFIs, it may also be necessary to first randomly select a subset of the MFI's operational regions. This can be done by assigning each region or branch a number and randomly sampling according to the relative size of each region (This technique is described in detail in section 2.4). For example, in one case study, the MFI operated in four regions, but because of time constraints and budget concerns, it was decided that survey localities could be randomly selected from only three of these four. Figure 2.1 illustrates the MFI's geographical levels used in sampling with clusters.

FIGURE 2.1 Example of sampling with clusters



2.2.2 Determine the stages of clustering required

For very small MFIs, one-stage clustering can be used where several geographical clusters are selected randomly from a list of all geographical clusters within the feasible survey area, and all qualifying MFI client households within the selected clusters are surveyed. However, larger or more dispersed MFIs will require at least a two-stage cluster approach. In addition to a random selection of approximately five to six geographical clusters, a second random sampling is done within each selected geographical cluster to randomly sample a subset of the client households located in each cluster area. Random sampling within a cluster usually requires a list of all new clients residing in the area, and a random selection method to sample households from this list.

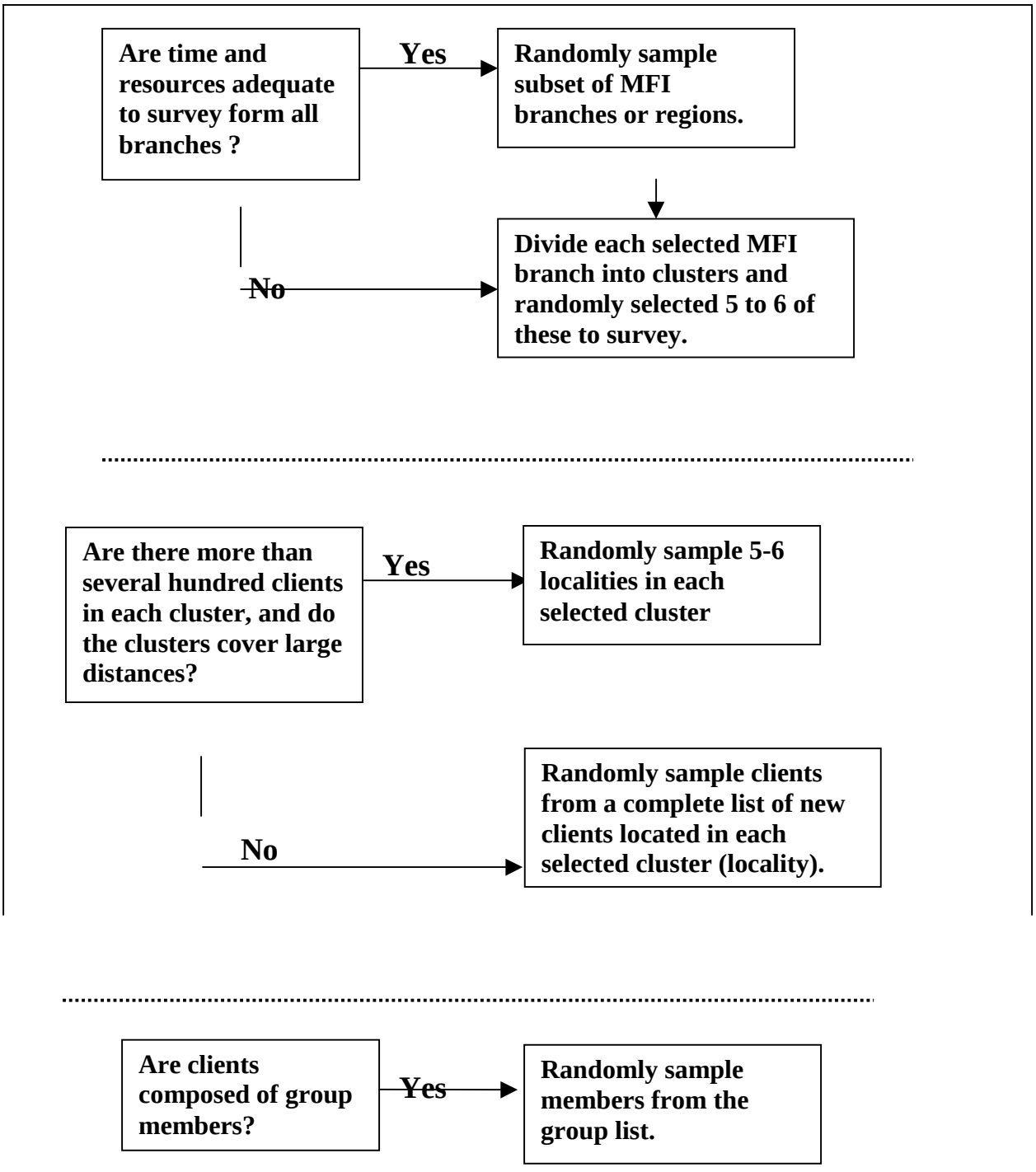
In some cases, a three-stage cluster may also be appropriate, particularly when MFI clients represent groups of individuals. Here random selection of client groups within each randomly selected geographical cluster would be followed by a random selection of members within that group. Figure 2.2 summarizes decisions used to determine the stages of cluster sampling to randomly sample new-client households.

Steps for sampling with clusters

1. Randomly sample a subset of MFI branches or regions if needed for larger MFIs.
2. Randomly sample a subset of clusters from a list of all MFI clusters in each region.

3. Within each selected cluster, develop a list of all MFI new-client households.
4. Choose a random-sampling technique to select client households to be interviewed.
5. If clients constitute groups of individuals, randomly sample groups within each selected cluster and then, for each group, randomly sample a subset of members.

FIGURE 2.2 Decision process for determining the stages of clustering



2.3 Step three: Determine the appropriate sample size

Calculating sample sizes is one of the most technically demanding aspects of survey design. On a practical level, sample size is partly determined by the time and resources available for the survey. On a technical level, four parameters enter into the decision what size of sample to choose:

- the desired precision,
- the probability distribution of the variable that the survey seeks to measure in the population,
- the choice of sampling design (whether single random sampling or multi-stage random sampling),
- the number of different variables (in our case poverty indicators) that we wish to capture.

Without a prior knowledge about the distribution of the poverty indicators among clients, a rule-of-thumb approach must be applied in determining the sampling size. Survey costs can be a major determinant. As a default, this manual recommends a sample size of at least 500, and maintaining a ratio of clients to nonclients in all survey clusters of 200 to 300, or 2 to 3. The larger sampling size for nonclients captures the presumably larger variance among nonclients with respect to any poverty indicator than the variance found among clients. Because of MFI targeting and self-selection of clients, the client group is likely less heterogeneous (has less variance) than the Nonclient group.

2.4 Step four: Distribute the samples proportionally

Ensuring that the sample selected represents the population of all MFI new clients is the main objective of the sampling process. Equal-probability sampling is one means of ensuring that each MFI new client has an equal chance of being selected. The equal-probability sampling can be applied two ways: probability proportionate to size sampling (PPS), and equal-proportion sampling (EPS).

The main deciding factor regarding which to use will depend largely on whether the research team can determine the number of new MFI clients in each cluster area. Often, the head offices of MFIs do not maintain detailed and accurate records of all new clients that include length of time in the MFI. This information is instead maintained at the branch level. If the research team can determine from the central office the number of new clients in each cluster, then the PPS method is appropriate. However, if the number of new clients in each cluster cannot be known without visiting each field office, then the EPS method may be logistically a more reasonable method. PPS is generally preferred over EPS because it permits equal numbers of clients to be surveyed in each sampled cluster whereas EPS requires that the number of clients in each survey locality be proportional to the total number of new clients located in all selected clusters.

2.4.1 Probability proportionate to size sampling (PPS)

PPS ensures equal-chance selection of households but requires that the number of new clients in each cluster be known before households are randomly selected. Of the two methods, PPS is the

easiest to implement in the field because the number of households surveyed are the same at each survey locality.

PPS is carried out in two stages. In the first stage, each cluster is assigned a chance of selection proportionate to the number of MFI new clients it contained, with the result that larger clusters have a better chance of selection than small ones.

In the second stage, the same number of MFI new-client households is selected from each selected cluster. Table 2.1 illustrates the steps involved in weighting the clusters proportionate to the size of new clients in each. Assuming only three clusters will be used for sampling, these are selected randomly using a random-number chart, where the range of numbers from 1 to 100 is assigned to each cluster according to the share of new clients each cluster holds: 1 to 16 for cluster 1, 17 to 37 for cluster 2, and so on. If the random numbers selected are, for example, 5, 18, and 60, then clusters 1, 2, and 4 are selected. Then an equal number of MFI new clients (67) is drawn from each cluster.

TABLE 2.1 PPS method of selecting MFI new client households

Cluster number	Number of new clients in cluster	Portion of MFI's total new clients	Sampling interval	Share of sample size	Sample size	
					Client households	Nonclient households
1*	600	0.16	1–16	0.33	67	100
2*	800	0.21	17–37	0.33	67	100
3	450	0.12	38–49	0	0	0
4*	1,250	0.33	50–82	0.33	66	100
5	700	0.18	83–100	0	0	0
Total	3,800	1.00		1.00	200	300

*Randomly selected clusters.

A number of software programs offer functions to generate random-number tables, such as the one in Figure 2.3.

FIGURE 2.3 An example of a random-number list

05	18	60	01	56	20	14	84	34	08
65	11	70	30	59	61	04	41	55	09
46	81	46	61	94	87	98	14	23	16
36	99	65	27	46	95	50	11	80	13
22	05	08	75	21	18	33	95	43	87
09	82	39	06	17	30	04	46	45	74
13	38	21	08	30	55	91	65	18	84
51	23	76	25	10	08	49	34	96	40
12	41	26	44	70	45	09	47	17	31
49	21	31	15	67	86	48	43	54	04

Example of using PPS for sampling

In Nicaragua, the new clients were randomly sampled from two large geographical areas on the basis of the number of new members in each area. Twenty percent of new clients were located in the northern region. Of the five branches randomly sampled, one was drawn from the north and four were drawn from the south.

Because the number of new clients was known for each area, and the sample frame was weighted to reflect the share of new clients in each area, the PPS method was used to determine the number of clients interviewed in each sampled branch. The list of clients was determined at the branch level by the credit agent at each branch. An equal number of clients (40) were randomly selected from each of the five branches. Nonclients were randomly sampled using the random walk at the actual survey site.

2.4.2 Equal-proportion sampling (EPS)

EPS attaches to each cluster an equal chance of selection regardless of size, but then distributes the numbers of clients interviewed in each cluster according to the share of new MFI clients in that cluster to the total number of new clients in the *selected clusters*. In this way, if new-client information is not centralized, evaluators need only determine the number of new clients in those clusters randomly selected. The evaluation team will eventually be required to visit branch offices to collect new-client information for that cluster.

The number of households interviewed in each cluster will differ. Bigger clusters will have more MFI households selected; and smaller clusters, fewer. Table 2.2 shows an example of how the EPS method can be applied.

2.4.3 The EPS method applied to client groups

In many MFIs, clients are members of financial groups. The individuals in these groups are each counted as a new client. However, when distributing samples within a cluster, evaluators will want to draw a sample of groups from which to randomly select households for interviewing. If the PPS method is used for sampling, the same number of households can be chosen from each group regardless of its membership size. However, if the EPS method is used, the number of individuals interviewed in each group needs to be adjusted according to what share of new clients are represented in that group to the total number of new members in the cluster.

Table 2.3 gives an example of how the number of members from each group is determined. In the example, assume that it is decided that the 46 interviews from cluster 1 are to be taken from 5 randomly sampled financial groups. The sample size for each group is adjusted according to the share of the groups membership to total membership in the cluster.

TABLE 2.2 EPS method applied to select households

Cluster No.	Probability of selection	No. of new clients	Weight	Sample size	
				Client households	Nonclient households
1*	0.20	600	0.23	46	69
2*	0.20	800	0.30	60	90
3	0.20	?	0	0	0
4*	0.20	1,250	0.47	94	141
5	0.20	?	0	0	0
Total	1.0	2,650*	1.00	200	300

*Randomly selected clusters.

TABLE 2.3 EPS method of selecting households in new client groups

Group number in cluster 1	Number of new clients in group	Share of total new clients in cluster	Sample size	
			Client households	Nonclient households
11	27	0.23	11	17
12	17	0.14	6	9
13	28	0.24	11	17
14	27	0.23	10	17
15	20	0.17	8	12
Total	119	1.00	46	72

2.5 Step five: Select the actual sample

The discussion so far has guided researchers in how to develop a sample frame using several levels of geographical clusters to systematically reduce the number of new-client households from which to choose a random sample. It has also guided researchers in how to determine the number of new-client households to be randomly sampled in each cluster such that each new-client household in the feasible area has equal chance of being selected. Once these two techniques are applied, the actual random selection of new-client households can proceed.

2.5.1 Random sampling within clusters

The research team will need to define how to randomly select clients within each cluster. A relatively straightforward method of selection is systematic sampling where draws are made at fixed intervals through a list of the sample population, starting from a random unit. The method requires that a list be made of all new clients or client groups within a selected cluster.

For example, suppose the survey team needs a sample of 10 from a list of 150 clients or client groups. First, a number is randomly selected between 1 and 15 (150 divided by 10) and, starting from a randomly selected client or client group on the list, every 15^{th} one is selected. If 5 were the

randomly selected number, then the sample would be composed of clients 5, 20, 35, 50, 65, 80, 95, 110, 125, and 140. A second method is systematic random sampling. Using this method, all new clients are assigned identification numbers and then selected either through selecting slips of paper with the numbers on them from a hat or through the use of a table of random numbers.

In addition to randomly sampling households to interview, the survey team should also prepare a second list of randomly sampled households to place on a reserve list in case a sampled household does not qualify for being interviewed or is unable to be interviewed. As a rule of thumb, the reserve list should contain an additional 4 reserve names for each 10 sampled names. Once the survey is underway, the first name on the reserve list is taken to replace the first sampled household falling out of the survey. All additional replacements are made in the order they appear on the reserve list.

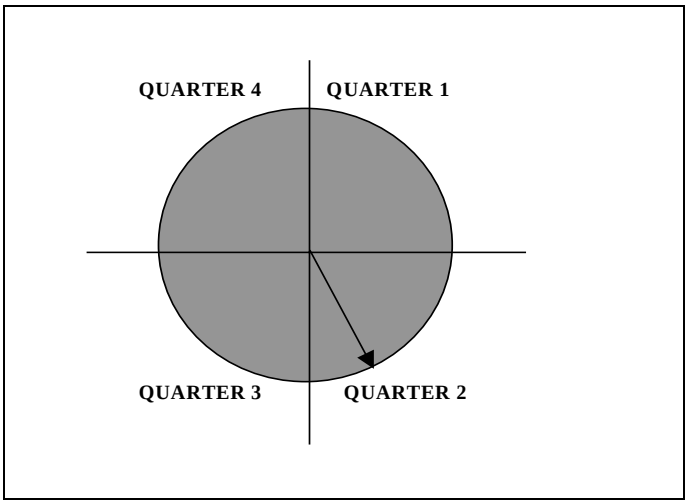
2.5.2 Random sampling Nonclient households: The random walk

Sampling Nonclient households would be a time-consuming exercise if an accurate list of all households had to be created within each survey area. The researcher can avoid this task by employing a two-stage technique called the EPI Cluster Survey Design method, or EPI method. Although the method may be less precise than sampling from a true population list, we think its greater efficiency makes it an appropriate trade-off. In contrast to the sampling frame for client households, the EPI method requires no preparatory work other than defining the boundaries of each survey site. The random selection of Nonclient households is done at the same time the survey is conducted.

The EPI method, developed by the United Nations Children's Fund to monitor the immunization of children within large areas, can be easily adapted to fit the MFI situation. The method is used within the local community or subdivision where sampled MFI clients live. The boundaries of the area can be set by asking MFI clients to identify landmarks in all directions that establish an outer perimeter of where client households are located.

As illustrated in Figure 2.4, the survey team identifies a central point in this area from which to divide the area into quarters (the area may be divided into more than four quarters if it is a very large zone). From this central point, the survey team selects a random direction by spinning a bottle or pen on a flat surface and noting the direction in which it points. The team selects only households lying in this direction or quartile. If desired, a second spin can be made to select a second direction.

FIGURE 2.4 Quartiles of a survey area



Depending on the density of households and the approximate area in which MFI clients reside, the survey team determines an interval number for selecting (sampling) and interviewing Nonclient households. In dense urban areas, an appropriate interval may be 10, so that every 10^{th} dwelling counted from the center located in the randomly selected direction would be sampled and interviewed. For rural areas, a much smaller interval number may be more appropriate. Interviewers may need reminding that households do not necessarily live in separate homesteads but can also live in housing complexes. Within a single building, a random process should also be defined for selecting which household to interview. Several households may be located within the same building, and renting and squatter households are also counted. Figure 2.5 outlines the random-walk process.

FIGURE 2.5 The random walk

1. Approximate the village or locality boundaries of sampled MFI new-client households and draw a rough map.
2. Determine a central point and assess density of households.
3. Divide area into quarters.
4. Randomly select one or more directions by spinning a pen or bottle to determine the quarter to be sampled. If households are dispersed, count all households within a quarter; if concentrated, narrow the count to a particular route within the quarter.
5. Follow the selected direction and select households in intervals of a predetermined number based on the population density (for example, every fifth household).
6. Replace dropout households by sampling the very next household.

2.6 Summary

Below is a brief summary of the steps to be followed in developing the sample design.

1. **Define the population and sampling unit.** The "population" refers to all clients and nonclients that reside within the MFI's areas of operation. The "basic sampling unit" is the household of new clients and nonclients.
2. **Construct the MFI-based sampling frame.** The sampling frame refers to how information on the number of new MFI clients within the MFI's operational area is used to determine which localities will be surveyed.
3. **Determine the appropriate sample size.** The minimum sample size is set at 500 households, of which 200 are new MFI clients and 300 are nonclients, or a 2 to 3 ratio.
4. **Distribute the sample proportionally.** Proportional sampling refers to techniques that structure the selection of households so that each has an equal chance of being selected.
5. **Select the actual sample.**

CHAPTER THREE:
ADAPTING THE POVERTY-ASSESSMENT QUESTIONNAIRE TO THE LOCAL SETTING

This manual provides a well-tested, fine-tuned list of questions that have been worded, coded, and ordered in a questionnaire format to produce consistent, measurable, and interpretable results. (See annex 3 for a copy of the recommended questionnaire). The core questions identified for the survey should be included and maintained in their general form under all circumstances. In combination, these questions build indicators that are later used for calculating a poverty index. The ways in which responses are grouped, sequenced, and measured are designed to support subsequent analysis of the survey data at a later stage.

The recommended questionnaire, however, does require some customization to fit local conditions. Evaluators carry the task of adapting the standardized questionnaire form to fit the national and, sometimes, localized setting. To avoid distortions that could weaken the reliability and validity of a question and its underlying indicator, those tasked to adapt the questionnaire can benefit greatly from an overview of the intended purpose of each section within the standardized questionnaire and a short summary of possible adaptations. The following sections outline the objectives and issues associated with each section and describe how each can be adapted without altering the underlying intent of the questionnaire.

3.1 Introducing the study and screening households

3.1.1 Introducing the study

Ideally, the introductory information is written ahead of time so that interviewers can introduce themselves and the reasons for the interview precisely and accurately to the household. The following points clarify the kinds of information provided to household respondents:

1. *Identify yourselves.* Households will be more cooperative if they know who is conducting the study. An important point to mention is that the survey team is not directly employed by the MFI. This disclosure will eliminate a potential source of bias if the household thinks its answers may affect its access to services from the MFI. If the survey team is associated with a well-known university or other institution, this may encourage cooperation and should be highlighted. On the other hand, if the organization is associated with certain parts of government, particularly local government, or a political party or ethnic group, many households may be reluctant to provide truthful information about their relative wealth or poverty. Downplaying these aspects may be prudent.

2. *Show letters of introduction and endorsement.* Most countries expect that outsiders first seek permission from local leaders before approaching households in a given locality. In addition to introducing the survey, these courtesy visits also can provide an opportunity to collect important information about the community being surveyed. In some cases, a letter of

introduction from MFI headquarters to MFI clients and by local authorities for nonclients can reassure the household and further facilitate introductions.

3. *Inform households of your purpose.* Most households will not fully understand the methodology used for this study. However, many will quickly fathom the overall purpose: to determine if the living standards of new MFI clients differ from nonclients living in the same area and, if so, in what ways. Further clarification of the purpose of determining whether the households of MFI clients are relatively poorer or wealthier is discouraged. This information could influence the way that questions are answered by the households and thereby introduce a major source of error in the results.

4. *Explain why the household has been selected.* Households also appreciate knowing that they have been selected for an interview on the basis of a random process. Those making introductions can draw analogies to such methods as pulling names from a hat to explain exactly what this means.

5. *Assure respondents of confidentiality.* In many countries, fear of crime or traditional beliefs may also inhibit many households from sharing private information. Introductions by the survey team should incorporate clear statements about the neutrality of the study team, and the confidentiality of information collected for the study. The study team should guarantee and subsequently follow through on their guarantee that no outside body will access the data for purposes other than those intended.

3.1.2 Screening households for applicability

Not all households qualify for participating in the study. After making introductions and before beginning the interview, the interviewer must verify that the household qualifies either as a new-client household or as a Nonclient household. A household identified as having a member who is a new MFI client can still be disqualified for two reasons:

- Someone else in the household is also a member of the MFI and has been so for longer than six months.
- Someone in the household was a member of the MFI but is no longer a member.

If a sampled MFI client household is disqualified, replace it with the next household named on the replacement list.

A household sampled as a Nonclient household can also be disqualified for the same two reasons:

- Someone in the household is a member of the MFI.
- Someone in the household was a member of the MFI but is no longer a member.

If either type of household is found, the interviewer should thank the members of the household for their time and terminate the interview. In the event that a nonclient household is disqualified, a replacement for it is the next household in the same direction.

3.1.3 Type of respondent and preferred venue for interview

In addition to verifying that the household is appropriate as either a new MFI client or Nonclient, determining who within the household responds to the questions is also important. Ideally, both the head of household and spouse of the head of household will respond. In many cases, if this is not possible, having either of these persons respond is the next best choice. The location of the interview will also partially determine results for several key indicators. Interviews ideally take place in the respondent's home, where the quality of housing and extent of durable assets can be observed.

3.2 The survey form

The following sections identifies specifically where within the questionnaire adaptations will be needed. Some changes will require altering the actual questionnaire form and others will require that a sheet of notes be developed that provide the definitions of question categories and terminology. Field staff will use the notes as a reference during the actual household survey.

3.2.1 Section A: Documenting households through identification information

Purpose

Figure 3.1 shows Part A of the questionnaire. Successful field surveys require adequate identification to distinguish among information from different households. Coding of households, client groups, localities, and clusters all are required to identify households in later stages of the analysis.

Issues

The range of numbers used by each interviewer for each household needs to be prespecified. We recommend that each type of identification variable be assigned a range of identity code numbers with a beginning and end point to eliminate any risk of overlap. Evaluators should assign identification code ranges so that they are easily understood. The types of codes that can be used are as follows:

(A2) Locality *codes*. Codes that link survey sites to government administrative localities are used to relate survey data to secondary data collected from other sources.

→Define: Assign numeric codes to each administrative locality in which survey sites are located and list on sheet of notes.

(A3) MFI cluster codes. Each questionnaire can also be partially identified by the MFI survey area in which it is located. A name and code number for each of these should be determined and written on questionnaires before interviews. The likely number of survey areas will be 5 to 6.

→Define: Assign numeric codes to each survey area and list on sheet of notes.

(A4) Group codes. If clients are organized into groups, then the name and an associated code number for the group become important identifiers.

→Define: Assign numeric codes to each group of clients surveyed and list on sheet of notes.

(A6) Household identification codes. In this study, the key means of identifying each household is through the assignment of a unique identity code. Given the sample size of 500, a household identity number can be three digits. These numbers can be assigned before the interview, or written on the questionnaire at the time of the interview. Non-overlapping household code ranges need to be assigned for each survey site. For instance, the first survey site could be assigned a range of 100 to 199; site two, 200 to 299; and so on.

→Define: Assign ranges of household identification codes for each survey site and list on sheet of notes.

(A11) Interviewer codes. Finally, to control for data errors and monitor interview performance, the questionnaire records the name and code of each interviewer as well as that of the person who has proofed the questionnaire in the field.

All coding associated with identifying households, members, areas and groups should be summarized on a sheet of paper and given to each interviewer to use as a reference. Table 3.1 shows the ranges of different types of identification codes used in a previous assessment survey.

TABLE 3.1 Example of questionnaire identification code ranges used to distinguish households

Type of identification code assigned in each cluster	MFI cluster area				
	1	**2**	**3**	**4**	**5**
Household	100–199	200–299	300–399	400–499	500–599
Client group	10–19	20–29	30–39	40–49	50–59
Locality	10–19	20–29	30–39	40–49	50–59

3.2.2 Section B: Family structure

Purpose

Characteristics—such as the number, age, health, education, and occupation of household members—represent indicators of the household's resources in the form of human capital. The purpose of this section is to quantify the key aspects of the household's investments in human capital. Specific objectives are as follows:

- To determine the composition of household members, based on survey definition of a household.
- To record selected poverty-related aspects of each individual member of a household.

FIGURE 3.1 Part A of the survey questionnaire

Assessing Living Standards of Households

International Food Policy Research Institute

A study sponsored by the Consultative Group to Assist the Poorest (CGAP)

Section A Household Identification

A1. Date (mm/dd/yyyy): __/__/____

A2. Locality code: ☐☐

A3. MFI cluster code: ☐☐

A4. Group code: ☐☐

A5.Group name: ☐

A6. Household code: ☐☐

A7. Household chosen as (1) client of MFI, or (2) nonclient of MFI? ☐

A8. Is household from replacement list? (0) No (1) Yes ☐

A9. If yes, the original household was (1) not found or (2) unwilling to answer, or (3) client status was wrongly classified: ☐

A10. Name of respondent:

Name of the household head: ☐

Address of the household: ☐

A11. Interviewer code: ☐☐ A12. Date checked by supervisor (mm/dd/yyyy): ___/___/____

A13. Supervisor signature: _____

Issues

Definitions of household vary widely. Working consistently from a standard definition is crucial for good measurement. For the purposes of this poverty-assessment method, a *household* is defined as a group of individuals who live under the same roof and regularly share meals and expenses together. A *family* does not necessary constitute a household, as it can include members who live away from home or who are closely related but do not cook together and pool resources to cover expenses. All household members should be screened to ensure that they fulfill all criteria set for inclusion as a household member. In some cases, the definition of a household may require inclusion of a head of household who works away from the home but contributes regularly to expenses and does not also

support any other household. Other family members living away from home are not counted unless they are children of the head of household attending a boarding school and the household supports them fully.

All qualified household members must be included in either table B1 or B2 of the survey form (Figures 3.2 and 3.3), and names as well as identification codes must be assigned. (These codes will be used again later in the questionnaire.) All columns in these tables represent poverty-sensitive aspects of individuals and, as categories, should not be changed. However, determining the appropriate wording for categories or responses may require some changes.

B1. HOUSEHOLD ADULTS

This part of the questionnaire is shown in Figure 3.2.

Adult ID code. Each member of the household receives a separate identification number. This number will be used consistently throughout the questionnaire.

Marital status of head of household. A household head can be either male or female. This category refers to his or her marital situation at the time of the interview. Note "a" below the table in Figure 3.2 lists the categories of marital status. These codes should not require editing.

Age. For older members, exact ages are sometimes not known. The respondent can be asked to estimate the approximate age of some adult members if finding out the exact age is likely to be time consuming.

Maximum level of schooling. This indicator can have sequenced responses coded so as to measure progressively higher levels of completion. The categories of response are listed in Note "d" below the table.

→Adapt: Identify the appropriate levels of educational advancement and list them progressively in terms of completion.

Can write. This refers to the ability to read and write, regardless of the language involved (all local languages apply).

Main occupation. This refers to the type of activity done most often by the household member on a daily basis. If individuals do more than one type of work, record the type that takes up the most time per day. If individuals spend the largest part of their day *not* working, it is critical to record this using one of the codes for not working. Categories of responses are listed in Note "f" below the table.

Amount of loans borrowed. This provides information on the extent to which the household has taken advantage of services from the MFI being assessed. Loans from all other MFIs are not included. This information will later help identify which households may have already benefited from MFI participation.

FIGURE 3.2 Part B1 of the survey questionnaire

B1. Adult members of household (aged 15 and above)

ID code	Name	Status of the head of the HH[a]	Relation to head of HH[b]	Sex[c]	Age	Max. level of school-ing[d]	Can write[e]	Main occupation, current year[f]	Current member of study MFI[e]	Amount of loan borrowed from study MFI	Clothes/Foot-wear expenses for the last 12 months in local currency[g]
1	(HH head)										
2											
3											
4											
5											
6											
7											
8											

[a](1) single; (2) married, with the spouse permanently present in the household; (3) married with the spouse migrant; (4) widow or widower; (5) divorced or separated; (6) living mostly away from home but contributing regularly to household.
[b](1) head of the household; (2) spouse; (3) son or daughter; (4) father or mother; (5) grandchild; (6) grandparents; (7) other relative; (8) other nonrelative.
[c](1) male; (2) female.
[d](1) less than primary 6; (2) some primary; (3) completed primary 6; (4) attended technical school; (5) attended secondary; (6) completed secondary; (7) attended college or university.
[e](0) no; (1) yes.
[f](1) self-employed in agriculture; (2) self-employed in nonfarm enterprise; (3) student; (4) casual worker; (5) salaried worker; (6) domestic worker; (7) unemployed, looking for a job; (8) unwilling to work or retired; (9) not able to work (handicapped).
[g]In order to get an accurate recall the clothes and footwear expenses for each adult are preferably asked in the presence of the spouse of the head of the household. If the clothes were sewn at home, provide costs of all materials (thread, fabric, buttons, needles).

Clothing and footwear expenses. Household expenditures on footwear and clothing can reflect the relative poverty or wealth of a household in many cultures. **Accuracy in measurement is critical, however, in ensuring the reliability of the variable because this indicator will be treated as the benchmark poverty indicator**. Expenditures are limited to those made by verified household members (not extended family members living and eating elsewhere) and do not include gifts to the household. Items given by one family member to another are also not counted as an expense. The amount of expenditure is the amount paid for the item at the time of purchase.

The time period covered is the past year, and most respondents will need to be provided reference points (sequence of notable holidays or time of year, such as Christmas, a family event or start of school year). Tailoring charges should be included. If items are made in the home, the costs of *all* materials used (for example, buttons, thread, and cloth) should be estimated for each person. The respondent should be encouraged to ask other household members if he or she is uncertain of the items and amounts.

B2. CHILDREN UNDER AGE 15

This part of the questionnaire is shown in Figure 3.3.

Characteristics related to children in the household are important indicators of relative poverty for many households. However, because many surveyed households have no children, the survey

questions are limited to documenting the number, age, and clothing expenditures on each child. The amount of expenditure for clothing and footwear is calculated in the same way described in the section on B1.

FIGURE 3.3 Part B2 of the survey questionnaire

B2. Children members of household (from 0 to 14 years)

ID code	Name	Age	Clothes/Footwear expenses for past 12 months, in local currency[a]
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

[a] Clothes and footwear expenses are asked for once those for adults have been recorded, and in the presence of the spouse of the head of the household. In case of ready-to-wear clothing and footwear items, include full price; in other cases, include cost of fabric cloth as well as tailoring and stitching charges.

3.2.3 Section C: Food-related indicators

Purpose

Household eating patterns are strong indicators of relative poverty and vulnerability.

Eating patterns can be affected by the relative poverty of a household in several ways. First, poorer households tend to consume foods on a less regular basis than wealthier households and may eat lesser quantities per person. In some cases, poorer households may skip meals or eat smaller quantities at meals, either during particular seasons or on a more regular basis. Second, poorer households tend to consume more of less costly foods and less of more costly foods. Third, poorer households often are less able to purchase staple foods in larger quantities at more favorable per-unit prices, or are less able to maintain stocks of either homegrown or purchased staples. Specific objectives to measure these aspects of food security are as follows:

- To document the quantity and frequency or regularity of food served by the household on a routine basis
- To identify and document consumption of specific foods that signal the spending power of the household
- To identify the degree to which households are able to purchase in bulk and maintain stocks of staple foods.

Figure 3.4 shows section C of the questionnaire.

FIGURE 3.4 Section C of the survey questionnaire

Section C. Food-Related Indicators

(Both the head of the household and his or her spouse should be present when answering for this section.)

C1. Did any special event occur in the last two days (for example, family event, guests invited)?

(0) No (1) Yes ☐

C2. If no, how many meals were served to the household members *during the last 2 days*? ☐

C3. If yes, how many meals were served to the household members *during the 2 days preceding the special event*? ☐

C4. Were there any special events in the last *seven days* (for example, family event, guests invited)? (0) No (1) Yes ☐

(If "Yes," the "last seven days" in C5 and C6 should refer to the week preceding the special event.)

C5. During the *last seven days*, for how many *days* were the following foods served in a main meal eaten by the household?

Luxury food	**Number of days served**
Luxury food 1	
Luxury food 2	
Luxury food 3	

C6. During the *last seven days*, for how many *days* did a main meal consist of an inferior food only? ☐

C7. During the *last 30 days*, for how many days did your household not have enough to eat everyday? (0) No (1) Yes ☐

C 8. During the *last 12 months*, for how many months did your household have at least one day without enough to eat? (0) No (1) Yes ☐

C9. How often do you purchase the following?

Staple	**Frequency served**
Staple 1	
Staple 2	
Staple 3	

(1) Daily (2) Twice a week (3) Weekly (4) Fortnightly (5) Monthly (6) Less frequently than a month ☐

C10. For how many weeks do you have a stock of *local staples* in your house? ☐

C11. If your household earnings increased by (US$10–$20), how much of that would you spend on purchasing additional food? (Estimate amount as 5% of GDP per capita.) ☐☐☐

(*Note:* Does not include alcohol and tobacco.)

Issues

The potential for bias in measuring food consumption is high, and several steps are needed to limit the chance of error. First, the recall period for recording food consumption must be kept short. Few individuals can remember what was eaten more than a week in the past. Second, food consumption patterns can be drastically altered during special events so that the occurrence of these must be controlled for in the questionnaire. In the questionnaire, all households noting special events in the past few days are asked to recall the period before the event. Third, the wording of food-related questions must be precisely stated and rigorously followed. Whether a meal is *served* or *prepared* can be interpreted differently. Some households cook only once per day but prepare enough food to serve at two meals.

(C1) Special event. The first question in this section screens for special events that may have disrupted normal eating habits during the past two days. If these occurred, respondents are asked to reply to eating practices before the special event. In the case of a special event, the respondent skips to the next question and resumes the interview (C3). The same screening for special events is repeated for questions related to consumption of luxury and inferior foods (C4).

(C2-C3) Number of meals served to household members. This is usually a reflection of local eating habits. If households usually serve three meals each day, the expected number would be three. If a morning meal is unusual, the expected number may be only two per day. It is crucial, however, that all interviewers use the same interpretation of what constitutes a "meal."

→Define: Standardize definition of what constitutes a meal and add to notes.

(C5) Luxury foods. The questionnaire requires that, for each survey, three foods be identified that are locally considered of high quality and relatively expensive for the average household, and that the frequency, *in days*, that each of these was served be recorded. Luxury foods are very specific to local climates and customs. Usually, luxury foods cannot be consumed regularly by all households but are more frequently consumed in wealthier households than in poorer households. Their consumption is also not restricted to special religious periods or cultural traditions. Meat, eggs, or dairy products—and some processed foods or sweets—can act as luxury foods in many parts of the world. In some cases, rice in a nonrice-growing region or wheat products in a nonwheat-growing region can be treated as luxury foods. Sometimes, what is considered a luxury food changes by seasons. For example, rice may be considered a luxury food during the maize-dominated agricultural season but may lose its luxury status during the immediate rice postharvest season when the rice price falls. Seasonality in price fluctuations should therefore be taken into account when determining what food groups are to be considered as luxury foods at the time of the interview.

→Adapt: Identify three foods that can be considered luxury foods, such as in the list below.

Luxury foods used in test countries	
Kenya:	meats, rice, wheat products
India:	lentils, dairy products
Madagascar:	meat, legumes
Nicaragua:	beef, poultry, cheese

(C6) Inferior food. An inferior food is the opposite of a luxury food. A clear signal that a food is inferior is when many households in a given area tend to avoid its consumption if they can afford an alternative to it. An inferior food is usually a cheap substitute for a standard staple, or a cheap item or dish to be served with a standard staple. Cassava has constituted an inferior food for many households based in rice-growing areas, where rice is the preferred staple.

→Adapt: Identify the food that can best be regarded as inferior in all survey areas.

Inferior foods used in test countries	
Kenya:	cassava
India:	coarse bread and chili
Madagascar:	cassava, yam
Nicaragua:	tortillas

(C7) Number of days not enough to eat in past month. This question is intended to measure short-term or seasonal food shortages within the household. It can be interpreted to the household as any condition where meals were either skipped because of a shortage of food or if members did not eat as much as they needed to feel full (went hungry for part of or all of the day).

(C8) Number of months with at least one day not enough to eat in past year. This question is intended to measure longer-term food shortages within the household and is used to balance any seasonal bias that may have entered into the previous question. If the household experienced food shortages in a previous season, this question will tend to capture it. Because the recall period is long, the interviewer may need to probe the household to recall particular seasons when food prices tend to rise during the past year.

(C9) Purchases of staple foods. This question measures how often households purchase staple foods. Lower-income households tend to purchase smaller quantities more frequently despite the higher associated cost, because of limited available cash.

→Adapt: Identify three staple foods or storable foods that are regularly consumed in the local area. Order responses from highest to lowest frequency.

Staple foods used in test countries	
Kenya:	maize, maize meal, potatoes
India:	rice, cooking oil
Madagascar:	rice, cooking oil
Nicaragua:	maize flour, beans, rice

(C10) Stock of storable staple. This question acts as an indicator of the household's ability to maintain a stock of a storable staple over an extended period, suggesting a higher level of food security for the household. Poor households often have to sell food stocks when cash is short, or have difficulty acquiring food stocks because of lack of funds.

→Adapt: Identify a staple that is frequently stored by households for longer periods of time.

Storable staples used in test countries	
Kenya:	maize
India:	pickled foods
Madagascar:	rice
Nicaragua:	rice, beans

(C11) Propensity to consume extra food. This question asks households to estimate how much of an additional monthly income would go towards purchasing additional food. The amount of money asked should be small enough to increase monthly cash flow by a portion of what they already have. We estimate this to be no more than 5 percent of the country's gross domestic product (GDP) per capita. For example, if the average GDP per capita in a country is estimated at $300[2], then 5 percent of this would be $15. This amount needs to be translated into local currency.

→Adapt: Calculate 5 percent of local per capita GDP, and translate this amount into local currency.

3.2.4 Section D: Dwelling-related indicators

Purpose

The quality of housing is partially determined by the relative poverty of a household. Indicators of dwelling quality include not only the size of the house but also the durability of materials used in its construction and the extent to which it is kept in a good condition. Finally, indicators of the facilities associated with the housing, such as toilet facilities and access to drinking water, also can measure aspects of its quality. Specific measurement objectives are to assess the quality and size of the household dwelling relative to others within the local area, and to assess the quality of facilities available to and used by the household. Figure 3.5 shows this section of the questionnaire.

Issues

Specific characteristics of household dwellings will vary by locality and culture. In some cultures, household dwellings may be numerous but located within one central compound. In urban areas, dwellings may consist of rented rooms within a single structure. It is therefore essential to adjust questions in this section to reflect the circumstances of the survey households. The dwelling is defined as all enclosed living spaces used by the family on a routine basis. Building structures used primarily for storage or livestock are not considered part of the dwelling.

A second issue is the bias that occurs in areas where the level of local infrastructure precludes a household from accessing certain amenities. For example, even wealthy households will not use electricity in an area where electricity is not available. This potential bias will be balanced out in the study through area analysis and does not need to be accounted for in the questionnaire.

[2] Unless otherwise stated, all dollars are U.S. currency.

(D1) Ownership Status. Knowing the ownership status of the dwelling can greatly assist in interpreting other information about the household and characteristics of its dwelling. Whether the house is built on squatter land can be a strong indicator of its general insecurity and vulnerability.

FIGURE 3.5 Section D of the survey questionnaire

Section D. Dwelling-Related Indicators

(Information should be collected about the dwelling in which the family currently resides.)

D1. What is the ownership status of dwelling? (1) Owned (2) Given by relative or other to use

(3) Provided by government (4) Rented (5) Located on squatter land ☐

D2. How many rooms does the dwelling have? (Include detached rooms in same compound if same household.) ☐☐

D3. What type of roofing material is used in main house? (1) Tarpaulin, plastic sheets, or branches and twigs (2) Grass (3) Stone or slate (4) Iron sheets (5) Brick tiles (6) concrete ☐

D4. What type of exterior walls does the dwelling have? (1) Tarpaulin, plastic sheets, or branches and twigs (2) Mud walls (3) Iron sheets (4) Timber (5) Brick or stone with mud (6) Brick or stone with cement plaster ☐

D5. What type of flooring does the dwelling have? (1) Dirt (2) Wood (3) Cement (4) Cement with additional covering ☐

D6. What is the observed structural condition of main dwelling? (1) Seriously dilapidated (2) Need for major repairs (3) Sound structure ☐

D7. What is the electricity supply? (1) No connection (2) Shared connection (3) Own connection ☐

D8. What type of cooking fuel source primarily is used? (1) Dung (2) Collected wood (3) Purchased wood or sawdust (4) Charcoal (5) Kerosene (6) Gas (7) Electricity ☐

D9. What is the source of drinking water? (1) Rainwater,dam, pond, lake or river (2) Spring (3) Public well—open (4) Public well—sealed with pump (5) Well in residence yard (6) Piped public water (7) Bore hole in residence ☐

D10. What type of toilet facility is available? (1) Bush, field, or no facility (2) Shared pit toilet (3) Own pit toilet (4) Shared, ventilated, improved pit latrine (5) Own improved latrine (6) Flush toilet, shared or own ☐

(D2) Number of rooms. This is defined as number of rooms used by the household for its living quarters. The definition of what constitutes a room needs to be specified to fit local conditions. Detached buildings within the same compound that house different members of the household would be counted; however, storage sheds would not.

→Define: Standardize definition of what constitutes a "room" and add to sheet of notes.

> In Kenya, the definition of a "room" included not only the rooms located within the household's main building, but all detached living quarters used by individual members.

*(D3) Type of roofing material.*This question requires that the common types of materials used in roofing be identified, the categories for choices defined and sequenced in order from lowest to highest quality, durability or cost. Where households have more than one dwelling within a compound, roofing material refers to only that on the primary structure.

→Adapt: Determine categories for types of roofing and sequence these using code numbers from lowest to highest quality.

> In India, roofs made of impermanent materials were most common among poorer households while roofs made of cement or tiles were most prevalent among less poor households. In Kenya, metal sheets roofed nearly all houses regardless of the poverty level of the household.

(D4-D5) Type of exterior walls and floors. The choices of materials used for exterior walls and floors will differ by locality. However, the sequence of choices should reflect an improvement in quality, determined through either durability or cost. The highest code number should list the highest-quality building materials. Again, type of exterior walls and floors refers to only those on the primary dwelling structure.

→Adapt: Determine groupings for types of walls (and floors) and sequence categories (and code numbers) from lowest to highest quality.

> Because the coastal areas of Madagascar are prone to tropical storms, most houses are built with light, cheap materials so that they can be easily rebuilt. The types of walls and roofing do not indicate well differences in poverty levels among households.

(D6). Condition of dwelling structure. This question relies on the interviewer's subjective assessment and assumes that the interviewer is able to view the dwelling structure. To make the measurement consistent, interviewers should have a common understanding of what constitutes "dilapidated," "in need of repairs," and "in good condition." The condition should not depend on the dwelling size or quality of materials used, since these have already been measured.

→Define: Standardize interpretation of three levels of structural condition and add to sheet of notes.

> In Kenya, dilapidated was interpreted to mean the dwelling was structurally unsound, in need of repairs meant that parts of the dwelling were obviously in need of repair, and good condition meant that no obvious repairs were needed.

(D7) Electricity supply. The availability and means of delivery for electricity differ by location. The choices of response for this question may need to be altered to better reflect local circumstances. The

ordering of choices should range from lower to higher access and cost. Responses should also be recorded where no electricity is available.

→Adapt: Determine category choices of how households access electricity and sequence response choices (and code numbers) from lowest to highest cost.

> IN INDIA, A RECENT ASSESSMENT FOUND THAT ONLY 20 PERCENT OF THE POOREST HOUSEHOLDS HAD AN ELECTRICITY CONNECTION COMPARED WITH 80 PERCENT

(D8) Cooking fuel. The type of primary cooking fuel will reflect location-specific conditions. Bias introduced by area differences in fuel use is addressed during the analysis stage. The choices of fuel types can be standard for all survey areas and should be ordered to reflect the lowest to highest cost of fuel.

→Adapt: Determine types of cooking fuels and sequence categories (and code numbers) from lowest to highest cost.

> In rural Kenya, a significant indicator of household's relative poverty was whether the household purchased or collected fuel for cooking.

(D9) Source of drinking water. The source of drinking water will be determined by local conditions. In general, drinking water sourced from open bodies of water, including open wells, are of lower quality than drinking water accessed through closed systems, either public or private.

→Adapt: Determine the main sources of drinking water and sequence categories (and code numbers) from lowest to highest quality.

> In Madagascar and Kenya, poorer households were more likely to use open sources of water for drinking water.

(D10) Type of latrine. The type of latrine is a component of housing and can be partially determined by the relative poverty of the household. The categories of responses will need to reflect local practices, but the choices of structure should be ordered from lowest to highest quality or cost.

→Adapt: Determine main types of latrines and sequence categories (and code numbers) from lowest to highest quality.

> In India, even most of the least poor lack a latrine within the homestead; however, the likelihood of having one is much higher for these households (28 percent compared with less than 1 percent of the poorest households).

3.2.5 Section E: Other asset-based indicators

Purpose

Accumulation of assets is strongly influenced by household income levels. Poorer households use income to meet basic needs and have little extra to invest in durable assets. Measuring the value of certain types of consumer durable assets can signal differences in relative poverty, so that complete valuation of all household assets is not necessary. Specific objectives related to this section are to record the number of selected consumer assets owned by the household by asset type, and to assess the current market value of these selected assets. Figure 3.6 shows section E of the questionnaire.

Issues

Asking households to itemize and value their durable assets can be a sensitive topic. The list of assets is limited, as much as possible, to observable assets. The list of assets also excludes items that are part of a business owned by a household member where the asset can be considered inventory. Inventory can be items that were either purchased with the intent to be resold or to be used to make products to be sold. If a household owns a radio or refrigerator that is located at the business but is not for sale, this may be counted as a household asset. Finally, the value of an asset is the money the household could receive from selling it at the current time. Interviewers may need to probe to establish an accurate value by asking how old the item is and whether it is in good working order.

In the case where new clients may have already received a loan from the MFI, all questions related to assets need to be screened to eliminate any purchases that were made as a result of the loan. This is best done by asking what items were purchased from the MFI loan and to exclude them.

(E1) Size of landholdings. Landownership is a good indicator of wealth in many developing countries. The amount of land owned by a household refers to the size (acres, hectares or other measure) that is owned by all household members. The land may be broken into subcategories to represent differences in its use or relative value. If households are likely to know the value of their land at current market rates, *and* are likely to state the amount accurately, the value of landholdings can also be asked. Land values will differ by location and type. The interviewer may need to ask the household for the value of each plot.

→Adapt: Determine appropriate types of landholdings to measure—agricultural or nonagricultural, cultivated or uncultivated—with standardized definitions for each. Determine if land values can be accurately assessed and, if so, standardize guidelines for assessment. Use local measures of land area.

Categories for measuring land`	
Kenya:	Used for agricultural production, not used for agricultural production
Madagascar:	Irrigated, nonirrigated
India:	Irrigated, nonirrigated
Nicaragua:	Irrigated, nonirrigated

FIGURE 3.6 Section E of the survey questionnaire

Section E. Asset-Based Indicators

E1. Area of land owned: Agricultural _____ Nonagricultural _____

Value of land owned: Agricultural _____ Nonagricultural _____

E2. Number and value of selected assets owned by household. (Ask household to identify any assets purchased with MFI loan and eliminate these from the table below.)

Asset type and code	Number owned	Resale value at current market price
Livestock		
1. Cattle and buffalo		
2. Adult sheep, goats, and pigs		
3.Adult poultry and rabbits		
4. Horses and donkeys		
Transportation		
5. Cars		
6. Motorcycles		
7. Bicycles		
8. Other vehicles		
9. Carts		
Appliances and electronics		
10. Televisions		
11. Video cassette recorders		
12. Refrigerators		
13. Electric or gas cookers		
14. Washing machines		
15. Radios		
16. Fans		

E3. What is your overall assessment of the general wealth levels of MFI clients? (1) Poor (2) Average (3) Rich (4) Don't know MFI ☐

(E2) Value of assets.

(1-4) Livestock. In many countries animals can be important assets for the household. Sometimes, however, households may be reluctant to number or value their animals. Also, primarily urban areas are less likely to have large animal assets.

→Adapt: Determine which, if any, animals can indicate relative household wealth and include them in table E2. Include up to four categories.

In Kenya, most households were reluctant to count livestock because it is thought to bring bad luck. In India, where cattle are considered sacred and rarely killed, many cattle owned by a household have little monetary worth. The value of cows kept by the household can be more informative than the number.

(5-9)Transportation. In many countries, ownership of means of transportation can delineate differences in relative poverty. Bikes, motorcycles, and other motorized vehicles vary in degree of ownership from country to country. People in mountainous regions may own fewer bicycles; people in urbanized areas, relatively more.

→Adapt: Screen the list of transportation assets to include only those used in the areas surveyed.

Few household-owned vehicles or bicycles in Kenya but the latter were an indicator of relative wealth in India.

(10-16) Appliances. All major appliances and electronics are considered good for differentiating relative poverty levels. Not all will be appropriate for every survey area, so that the list should be amended to reflect only what is realistic.

→Adapt: Screen the list of appliances to include only those found in the areas surveyed. Choose questions about animals that are likely to be relevant for the majority of households and to be answered accurately.

In India, ownership of electric fans was a significant measure for signaling relative wealth. In the Kenyan highlands, few households own fans but many own televisions. In Nicaragua, most surveyed households owned at least one television, and its value was a significant determinant of the household's relative wealth.

3.3 Customizing the questionnaire using principles of questionnaire design

3.3.1 Adapting existing questions and identifying appropriate local indicators

The questionnaire for this study was tested in four different countries and found effective in measuring the relative poverty of households. However, the standardized form of the questionnaire will require certain adaptations to fit local circumstances.

Customizing the standardized questionnaire will be required for two reasons. First, in adapting the response categories to standardized questions, evaluators will need to reword some sentences. Second, evaluators may want to include location-specific indicators that capture a very specific

indicator of poverty. An example of a significant local indicator is measuring for remittances from overseas in Nicaragua. In either case, additions of local indicators must be made very sparingly. This tool limits the researcher to no more than *five* additional indicators to capture location-specific poverty measures.

3.3.2 Guidelines for writing well-worded questions

A questionnaire is a standardized form for collecting data from respondents for the purpose of measurement and quantitative analysis. It contains questions that are to be asked in the same way to all respondents, with answers recorded using standardized sets of response categories.

Questionnaire design is more of an art form than a scientific process. The quality of the questionnaire depends on skill and judgment, a clear concept of the information needed, how this information will be used, and an awareness of possible sensitivities of respondents. Good questionnaires are often developed in stages and involve extensive pretesting. Characteristics of good questions include:

1. The wording of a question determines whether the researcher and the respondent interpret the meaning of the question in the same way. No single wording of a question is correct. Instead it is important to understand clearly what effect a particular wording can have on the response.

2. The words used in a question should be familiar to the interviewer and respondent. They should correspond to local word usage and practices. They should not be ambiguous, with more than one possible meaning.

3. Leading questions tend to suggest to the respondent what the answer to the question should be. "Did you eat three meals today?" is more leading than "How many meals did you eat today?" Likewise, a biasing question includes words or phrases that indicate approval or disapproval. "Can you afford a telephone?" is more biasing than "Do you have a telephone?"

4. To avoid implicit alternatives, questions should state clearly all relevant alternatives to a question unless for some reason this is not appropriate. Also, in expressing alternatives, the order of presentation can affect responses, since those listed last tend to be chosen. To avoid unnecessary estimates, phrase questions in a way that allow the researcher to make calculations later. Asking the household how much sugar it consumed in the past week is easier and likely to be more accurate than asking how much it consumed in the past year.

5. A double-barreled question is worded in a way that respondents are required to give two answers when a place is available for only one response. For instance, asking in the same question how much rice and sugar the family consumed in the past week would be difficult to answer and interpret.

6. Finally, keep a clear frame of reference. Keep questions specific to the household being interviewed. Instead of asking whether households in the area had enough food to eat in the past month, which elicits an opinion or subjective assessment, it is more accurate to ask only whether the household being interviewed had enough to eat.

The respondent must be able and willing to answer the question in the way it is posed. The respondent may have trouble answering some of the questions either because he or she is uninformed or because he or she is forgetful. In this case, it may be necessary to involve other household members in determining the correct answer. Unwillingness to answer a question can manifest itself when either the respondent refuses to provide an answer or the respondent purposely provides a wrong answer. The likelihood of this occurring can be reduced if respondents have a positive perception of the interviewer and consider the information needed for a legitimate purpose. Providing a good introduction and not rushing the interview can help to relax the respondent.

Well-posed questions

- use simple and clear words,
- are not leading and avoid biasing the answer,
- avoid implicit alternatives and assumptions,
- avoid estimates,
- are not double barreled,
- consider the frame of reference,
- indicate to which respondents the question applies, and
- are not rushed.

3.4 Precoding the questionnaire

The standardized questionnaire developed for this manual is formatted to provide precoded responses to qualitative questions. Coding refers to assigning numbers to each response or category of responses for a given question. Qualitative questions on the questionnaire form are closed, that is, categories of responses are identified and a number for each is given on the questionnaire. Codes for different responses or categories of responses have been structured so that all possible responses can be categorized into one of the predetermined choices, and that no overlap exists between response categories and codes. In addtion, categories of responses are sequenced from lowest to highest quality or cost to better support data analysis. These principles should be adhered to as questions are adapted to fit local conditions.

Precoding questionnaires greatly accelerates the process of entering data into the computer and later analyzing it with statistical procedures. Computer-based statistical analysis requires codes in a numeric form. The questionnaire provides a code box for each question. The list of response categories and their associated codes for text questions are located after or below the actual question.

If the coding of a question needs to be adapted for local conditions, the following rules must be adhered to at all times:

- Coded responses are numbered and presented sequentially.
- Coded responses are located as close to the question as possible, and placement is consistent throughout the questionnaire (always at right, or always at left).
- Code boxes for each question are easily distinguishable so that interviewers are not confused with which one to use for a given question.
- Code responses are all-inclusive, so that the interviewer will not need to write in a response that does not fit into the categories provided.
- Inclusion of codes for "other," "don't know," or "not applicable" are avoided as much as possible.
- Quantitative responses are usually not coded. Recording the actual age of an individual is preferable to assigning codes to age groups and recording these.
- Yes or no responses are coded with no as "0" and yes as "1."
- The number of code boxes provided for each question matches the highest number of places possible for an answer. For instance, a coded question with more than nine response codes requires two boxes.
- The choice of codes for table-formatted questions is located below the table, but always on the same page.

Coded responses need to be clearly defined for the interviewer. In addition to a brief description of each code choice on the questionnaire, a separate sheet is usually needed to define exactly what each category includes. This separate sheet provides the interviewer with local translations for terminology, and the meaning of categories of responses in local languages. In addition, the definitions decided for what constitutes a household, or how a room is defined, can also be written out. The sheet should order information in the way it is presented in the questionnaire and give the reference number to each question.

Precoding is the responsibility of those adapting the standard questionnaire to local conditions and is best done at the same time each question is adapted. Identifying meaningful, all-encompassing yet nonoverlapping response categories often involves considerable reflection, and the task should not be left to junior members of the evaluation team.

CHAPTER FOUR:
PLANNING AND ORGANIZING ASSESSMENT IMPLEMENTATION

4.1 Guidelines for contracting the assessment

Selection of qualified local researchers is critical to the success of the poverty assessment. The ideal local researchers must have at least several years experience in conducting statistical socioeconomic household surveys in the operational area of the MFI and in supervising data entry, data cleaning, and tabulation. These researchers also have a track record for successfully completing research projects on time and budget. The researcher can also provide proof of his or her past experience in conducting data analysis for publication. Finally, both the researcher and institution for which he or she works should be reputed and neutral politically, ethnically, and religiously. The researcher should be prepared to work independently and stay above influence from other sources interested in the assessment's outcome.

The choice of the local researcher, whether with the institution or as an individual consultant, should be based on the experience, availability, and cost of a principal individual, and this person's participation should be tied to the actual project assignment.

The research agreement between the donor and the local researcher fixes the term of the agreement, status of the researcher, scope of work, responsibilities of the parties, payments, reports and delivery schedule, copyright and ownership of the work, and cases of dispute and termination. Regarding the scope of work, the following points can be specified:

- First installment of funds paid to researcher. Amount approximately one-third of field operations budget.

    1. The researcher compiles information and data to set up a sampling frame for selection of branch offices for the MFI and identifies experts/gathers data to assess poverty levels regionally by month 1.
    2. The researcher adapts standard questionnaire to fit local conditions and identifies up to five additional local indicators, if deemed necessary.
    3. During a meeting with the contracting party, scheduled during month 1, the researcher finalizes the questionnaire, trains enumerators, and randomly selects of survey clusters. Agreement is reached on how to randomly select clients and nonclients.

- Second installment of funds delivered. (Remaining two-thirds of field operations budget)

    4. The researcher implements the survey by end of month 2.
    5. The researcher finalizes cleaned data at the end of month 2, or beginning of month 3 and delivers to contracting party.

6. The researcher analyzes the data and computes a composite poverty index by end of month 3.
7. The researcher conducts expert interviews and or analyzes regional poverty data to assess general poverty level in survey areas to national levels. Calculates regional ratio.
8. The researcher writes a draft report containing the data and index described in point 6, and comparative ratios by month 4.
9. The researcher organizes and participates in a seminar at which the methodology and results of the poverty assessment are presented.
10. The researcher submits to contracting party a final report by end of month 4.

- Final payment of researcher fee and any overhead charges upon delivery and acceptance of final report.

4.2 Determining the time frame required for the study

The implementation period of assessment tool refers to the period from when the first decisions are made to conduct the study to the completion and distribution of final results.

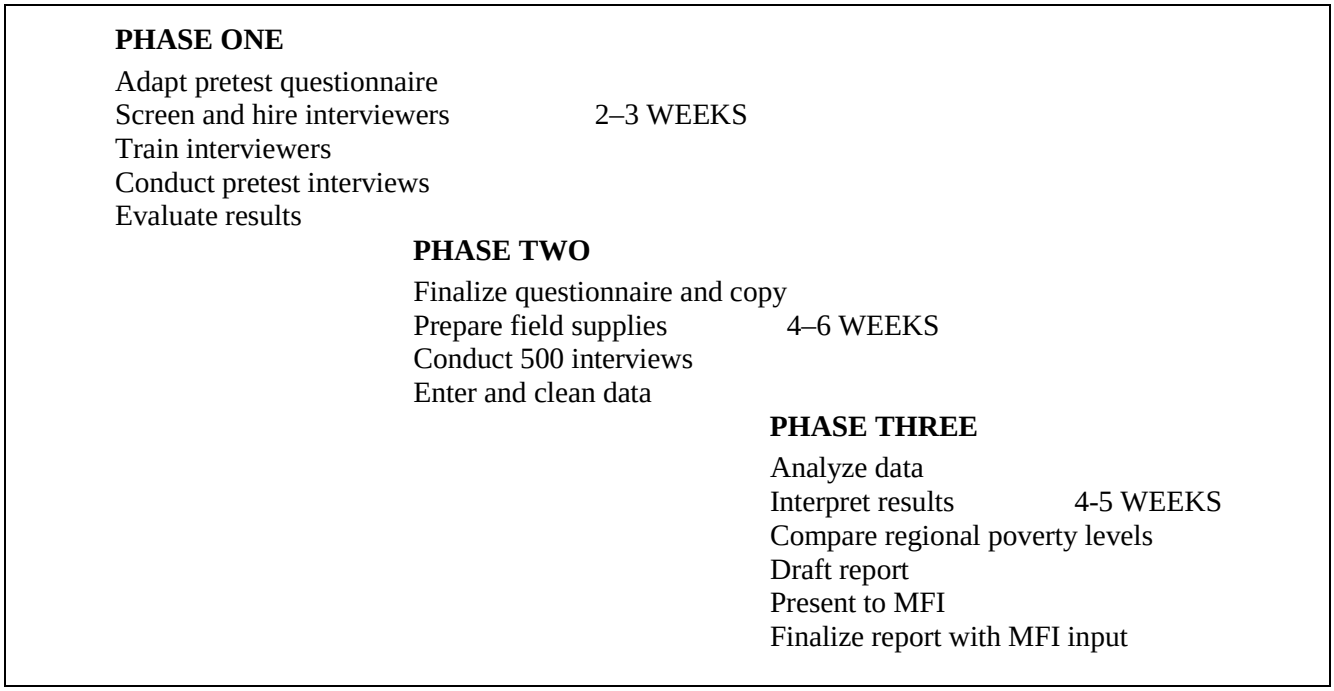
A time frame for starting and completing the study, including the sequencing of various activities, needs to be set early in the process. The time needed and the amount of overlap between activities should be carefully estimated, and the evaluators should be careful not to cut corners to save time. Also, field operations are best scheduled to avoid major national or religious holidays or periods of bad weather or heavy workloads. Figure 4.1 provides a list of activities and estimated time frames for the MFI poverty assessment. The estimates of time allocation are based on what was actually used in the four country case studies carried out to test the assessment tool. We estimate that the assessment can be completed in approximately four months.

In general, field operations are the most expensive stage of this methodology. Planning and managing well the time needed will contribute greatly to holding down costs. Successful field operations have five aspects:

- Time schedule,
- Budget,
- Personnel,
- Logistical support, and
- Performance measurement.

The quality of field survey implementation can make or break a study. Planning ahead the actual implementation is a critical step to preparing for successful fieldwork.

FIGURE 4.1 Time allocation for study by phases of activities



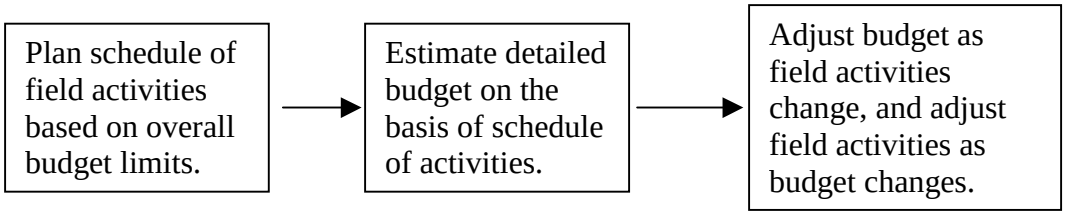
Survey oversight, data analysis and preparation of a final report is estimated to involve approximately 4 to 6 weeks of effort for a trained and experienced professional working with a cleaned electronic data set. This work will include the following major tasks:

1. Coordination of field survey to adapt questionnaire and test questionnaire;
2. Estimation of the poverty index and calculation of ratios, following methodology presented in the manual, and preparation of all statistical tables;
3. Qualitative and quantitative assessment of poverty levels in survey areas in relation to national averages;
4. Meeting with MFI staff to present results and gather feedback for any needed changes;
5. Preparation of final report.

4.3 Allocating budget for the assessment

4.3.1 Costs and planning associated with the field survey

Estimating the budget needed for the field survey requires careful scrutiny of how the field survey will be carried out. The allocation of the budget follows closely the schedule of activities.



Effective budgeting and cost control requires a detailed breakdown of the major cost categories that correspond to specific survey activities. The major expenses incurred in the MFI field survey will be personnel wages and per diems, transportation, fuel and related costs, and reproduction of questionnaires. Additional expenses may include office and computer rental, and telephone and other communication costs. Survey personnel need to review the budget regularly to ensure that the cost estimates remain in line with actual field progress. A small contingency fund is also needed to cover unforeseen expenses.

The actual cost of implementing a field survey will vary depending on the country or region in which the survey is conducted and on the rates charged by the contracting institution. In case studies, actual field costs ranged from a low of $4,000 to a high of $16,000. Figure 4.2 is a budget outline for an MFI study located in a rural area. It is estimated that the average cost for the field survey will be near $10,000. An additional fee for the data analysis and report preparation will average between $2,000 and $3,000.

4.4 Personnel, logistical, and performance issues affecting field implementation

Skilled personnel who are well trained and motivated can strongly influence the success of the field operation. A project manager will take overall responsibility for planning and implementing the field survey. The manager may also be the primary researcher for the project or may work with the researcher or researchers in coordinating the field survey. Ideally, the project manager will have previous survey experience and a good track record for successfully managing resources and personnel.

Field staff with prior field-survey experience are also desirable, but just as important are individuals with strong communication skills who can carry out interviews in a confident and relaxed manner while maintaining their train of thought. All interviewers require thorough training that includes an in-depth review of the questionnaire to understand its intent, and repeated practice in posing the questions in the local language. The project manager, supervisors, and interviewers need to participate in the training to ensure that all maintain a common understanding of how to use the questionnaire.

In addition to interviewers, the study will require at least one field supervisor, and if more than one team is planned for fieldwork, then several trained supervisors, each to head a team of interviewers. It is recommended that supervisors manage between 3 and 6 interviewers. The supervisors assign clear lines of responsibility and accountability to control the time frame and field costs. Supervisors also take responsibility for ensuring that the questionnaires are filled out correctly and completely and that the information contained in them is accurate before leaving each survey area. The supervisor makes random visits to a few households at each survey site to check the accuracy of the

answers. This spot-checking has an important quality-control effect.

FIGURE 4.2 Example of a survey budget for poverty assessment

1. Workplan and Personnel Required		
	Interviews	**Days**
Sample size (number of households)	500	
Interviewers	6	
Households interviewed per interviewer	84	
Households interviewed per interviewer per day (estimated)	5	
Field days per interviewer (including site identification and pretesting)		**22**

2. Personnel Costs

Personnel	**No.**	**Per diem rate ($)**	**Salary rate ($)**	**Number of field days**	**Total ($)**
Training and pretest					
Project manager	1	12	16	4	112
Field supervisors	2	10	10	4	160
Interviewers	6	10	8	4	384
Driver	2	8	6	4	112
Subtotal					**$768**
Data collection					
Project manager	1	12	16	5	140
Field supervisors	2	10	10	18	720
Evaluators	6	10	8	18	1,944
Driver	2	8	6	18	504
Subtotal					**$4,076**

Data management	**No.**	**Salary rate**	**Total ($)**	**Subtotal**
Sampling frame	4	15	60	
Data preparation	8	8	64	
Data entry	8	8	64	
Data cleaning	4	15	60	
Subtotal			**228**	**$228**
Total personnel costs ……				**$4,304**

3. Transport costs

Distances within 150 kilometers of central point	**Number of vehicles**	**Kilometers to site (return)**
For site identification	2	300
For pretest	2	300
For data collection	2	900
Total kilometers per site	2	1,500
Total kilometers traveled within the 5 districts (1,500 × 5)		**7,500**

Vehicle maintenance and fuel costs	**Estimated rates ($ per km)**	**Estimated cost ($)**	**Subtotal ($)**
Vehicle maintenance for 7,500 kilometers	$0.10	750	
Fuel for 7,500 kilometers	$0.20	1,500	
Subtotal ($)			**$2,250**

Total budget ($):

1. Personnel costs (from [2])	4,304
2. Transport costs from [3])	2,250
3. Other additional costs	
Photocopy/supply costs	500
Office, software	600
Data analysis & report preparation	2,000
4. **Subtotal**	9,654
Overhead costs (10% of 4)	965
Total cost	**$10,619**

In many cases, personnel involved in the field operations may be the same as those who later participate in data analysis. In past cases, field supervisors also worked as research assistants, and interviewers entered and cleaned data. In these cases, the individuals involved had previous experience doing both types of tasks.

A Kenya example of field implementation

The field personnel for a survey in Kenya consisted of one manager, two field supervisors, six interviewers, and two drivers. The field staff was split into two teams. Each team traveled to a different survey site. The teams needed a day to travel to survey sites, find accommodation, locate the homes of sampled clients, determine the boundaries and direction for the random-walk sampling of Nonclient households, and, where possible, line up interviews for the following day. Once settled, the interviewers were able to interview an average of six to seven households per day. Counting the day of preparation, interviewers were targeted to complete an average of five interviews per day. More than this was considered infeasible, and much less was considered avoidable with good organization. Time in the field totaled for both survey teams totaled 22 days, with interviewers working Saturdays but not Sundays, and avoiding interviews at night.

In addition to the field team, several data-entry specialists began entering data once interviews from the first two survey sites had been completed. In this way, the data entry was complete one week after the survey team returned from the field. There was a short overlap for interviewers to help with cleaning the data.

Well-planned logistical support—coordination around transportation, communication, field supplies, and contingency plans for disruptions—also greatly enhances the quality of field implementation. Logistical support needs to be carefully planned at all stages of the survey, especially where operations take place in remote locations with limited infrastructure. The right vehicles are large enough to carry the field team and supplies and sturdy enough to withstand the road conditions in survey areas. The estimated time needed for moving from site to site is based on a careful review of distances and road conditions. Communication methods and emergency plans also should be

identified beforehand. Access to petrol stations, food, and accommodation also need to be determined. Finally, those planning the logistics should consider local customs and political circumstances to avoid an unfriendly welcome or hasty exit.

Performance measures for field implementation are a means of controlling activities and meeting objectives. Performance measures need to be carefully thought through and well defined to ensure that they are not misunderstood or do not send unintended incentives. The most common measurement is the number of complete interviews to be finished each day. The actual interview is estimated to take only 20 minutes, however, locating households, making introductions and smooth departures can easily double the time needed. A target *average* number of five per day for each interviewer is recommended; however, some adjustment may be needed to reflect survey conditions. More than this could compromise the quality of interviews and less than this could increase the costs associated with time in the field. Expenditure limits can also be an effective means of measuring performance.

4.5 Training needs for survey interviewers

Interviewer training is best done by the same individuals who revise and finalize the questionnaire, and, ideally, who also undertake the data analysis. This method will avoid any confusion on the intent of questions and how they are worded and translated.

Interviewer training can follow a progressive format aimed at building skills step by step. Normally, interview training takes a minimum of two to three days, with an additional two days required for pretesting. The major determinant of how long the training will take is the amount of field experience both the interviewers and field supervisors already have.

4.5.1 Stage one: Summarize the background, purpose, and methodology of the survey

Format: Brainstorming and roundtable discussion of topics.
Materials: Written summary of information, flip charts to jot comments, notebooks for participants.
Time: Full day

Interviewers are critical to collection of quality data and are the individuals who must convincingly present the study to respondents, guarantee that the wording of the question is followed, and clarify issues for confused or reluctant respondents. To do their jobs well, interviewers need a solid overview of the following:

- Purpose of the study,
- Sampling frame to be used for identifying households,
- Field operations plan,
- Role of the interviewer and principles of good interviewing,
- Potential sources of error.

An informal presentation with opportunity to ask questions is a good format for conducting this stage of training. Examples of approaches to building interviewer understanding of each subject to be covered in this session are provided below.

DISCUSS THE PURPOSE OF THE STUDY

This session can be started by presenting a summary of the study's purpose and describing the roles of various organizations involved in it. Once interviewers have a general idea of what the study is about, the trainers can ensure that the interviewers have a more thorough understanding of the purpose and importance of the field survey by asking them probing questions that will urge them to think about why and how the study is being done. Chapter 1 describes the purpose of the study. The following are possible questions for spurring discussion:

Q: *What do we mean by poverty?*

A: Most countries have very clear definitions of what poverty means, and calculating a household's poverty level is very complicated. Poverty is usually measured in absolute terms, providing a measurement of income or expenditures in current terms to determine whether the household is poor. Collectively, the poverty line in a country is the cutoff annual income below which households are considered poor.

Q: *What is the difference between absolute poverty and relative poverty?*

A: This poverty-assessment study does not measure actual household poverty levels because it would be too involved for the scope of the study. Instead the study compares one household's poverty level with other households living in the same area. The interviewers in the study do this by asking questions that indicate how wealthy or poor a household is likely to be.

Q: *What kinds of household characteristics provide clues to the household's relative poverty level?*

A: Far too many characteristics exist to ask about all of these. We have tested many different questions and have concluded that the most informative are ones related to the kinds of resources the household has either in the form of assets they own, or in the kinds of occupations or the level of education they have. Other characteristics relate to how well they can meet their daily needs, so we ask about the quality and adequacy of their food, water, clothing, and housing.

In addition, questions should be focused on characteristics that can be observed, are not too sensitive to the respondent, do not involve too much calculation, and that will be answered accurately.

Q: *Why are MFI clients compared with nonclients?*

A: Many MFIs want to reach the poorer households in their areas, but determining whether the MFI is actually doing this is difficult. Outside donors often fund MFIs with the understanding that they are reaching the poor. This study is conducted to see how well MFIs actually do reach poorer households in the areas where they operate. We interview Nonclient households to determine how MFI clients compare with the general population in their operational area.

Q: *What motivations could MFIs have to influence the results of this study, and how might they try to do this?*

A: MFIs may want to have many of their clients perceived as poor, since this will result in a more favorable image of the MFI among donors and the local population. A few MFIs may advise their clients to underestimate their assets, the quality of their food, or other poverty-assessment factors. The best way to guard against this is for the interviewers to not reveal to MFI staff in detail the kinds of questions being asked.

DISCUSS THE SAMPLING FRAME USED FOR IDENTIFYING HOUSEHOLDS

Interviewers need to appreciate why interviewing the actual households sampled for this survey is so important. Client households are sampled from lists and replacements taken only from reserve lists. Discuss how these lists are created and how to decide when a replacement is needed. Describe your plan of how interviewers communicate with the supervisor to identify a replacement client household on the list if needed.

Sampling Nonclient households will require interviewers to understand thoroughly the random walk process involved in case interviewers need to find these households without a supervisor available to help them. Present the material from section 2.5.2 as simply as possible. Test their understanding by asking how the sampling might be done in different settings and what problems they are likely to face. Talk about how these problems can be solved:

- *Choice of houses to count.* When a direction has been chosen at random, a house out of a predetermined interval-number of houses will be selected to conduct the survey. The interval number is fixed according to the size of the area and the number of households that will be surveyed so that there is a large coverage of the area. Generally, one house out of 5 can be chosen in a sparsely populated area, one house out of 10 or 15 in larger or more denselyt populated areas. All the houses must be carefully counted, even shanties or temporary structures; these are likely poorer households. Buildings that are not residential houses (for example, churches, schools, mosques, and city halls) are not counted.

  In the street, the houses are counted alternatively on the left side and on the right side of the street; where an intersection occurs, the enumerator will go alternatively on the right side and on the left side.

- *Multiple households in a house.* When two or three households live in the same building, a number is given to each household and one of them is selected at random (for example, writing the numbers on small piece of papers, and selecting at random one of them). When a building has a large number of households, each household is given a number and several households can be selected. For example, if one house out of 10 is selected and 30 households are in the building, 3 households are selected.

- *Random walk in a city.* For a small town (less than 4 to 5 kilometers from one side to the other), the enumerator must identify a central point in the town (such as a market place, and intersection of main roads, the city hall, or a main building) from where to begin the random walk.

  If the city is large, only the locality or ward where MFI sampled clients are concentrated is surveyed. Ideally, clients will live close by so that boundaries can be determined. If no clear boundary can be defined, a rough map can be drawn with the help of local MFI clients or staff. The map must show natural boundaries (such as rivers, main roads, or parks) so that

different parts can be identified. A number is given to each area and some areas are selected at random for the surveys. For example, 3 to 4 areas can be selected out of 10 areas in the city. In each area selected, a central point must be identified to begin the random walk.

- *Random walk in a rural area.* Distances are generally far larger in the rural areas, and in the developing countries, official maps are usually missing. In this case, the administrative area that will be surveyed can be divided into several zones on the basis of natural boundaries. Several zones are selected at random. For the choice of the houses, when no street is clearly marked, the interviewer must follow the direction taken at random.
- *Replacement households.* With the random-walk technique, the replacement household will be taken in the house ($n + 1$). For example, when one house out of 10 is selected and the household chosen doesn't want to answer the questionnaire, the 11^{th} house is chosen. When the households are absent, the interviewer should avoid taking the replacement household immediately. The household members may only be at work or momentarily absent. In this case, the enumerator should try to come back when the household members are present (during the evening or weekend). If absent households are always replaced, the sample could underrepresent working households, for example.

PRESENT YOUR FIELD IMPLEMENTATION PLAN

Interviewers will have many questions about the kind of support they will have in the field. They will want to know about vehicles, what kind of accommodation they can expect, when they will do the work, what kinds of materials will be provided, what expenses will be covered by per diem, and what measures will be taken if they are ill or become injured. Review the schedule you have set for implementing the field survey and the job responsibilities of each person involved in the field survey.

DEFINE THE ROLE OF THE INTERVIEWER AND REVIEW THE PRINCIPLES OF GOOD INTERVIEWING

The following are important points to highlight about the reviewer's role and the art of good interviewing:

- Interviewers should be open-minded about the timing of interviews for this survey and should be willing to schedule interviews at a time when respondents are able to meet with them. Interviewers should be forthcoming about the amount of time required for the survey—20 minutes.
- Interviews should be rehearsed to the point that the interviewers are able to ask questions using the precise wording written, while maintaining a relaxed and conversational voice. The questions should also flow from one to the next, without breaks while the interviewer finds his or her place, or reads through the next question.
- Interviewers should know the questions well enough to add clarification and encouragement if the respondent is confused or hesitant to answer. If an answer does not sound confident, interviewers should gently probe to verify that the answer is well thought through.
- Interviewers should also maintain a pleasant and clean appearance, and behave politely.

- Interviewers should maintain a neutral stance on the questions being asked, on the survey's purpose, and on the MFI being assessed.

DISCUSS MAJOR SOURCES OF ERROR IN THE FIELD AND HOW TO CONTROL FOR THESE

Inevitably, field staff will encounter difficulties in implementing the survey as planned. How well they handle these difficulties, however, can greatly reduce the likelihood of error in the data set. Some common sources of errors in the field are discussed below.

- *Sample selection errors.* Sample selection errors can come from several sources. First, there may be a temptation to exclude certain clients or locations for reasons that do not follow the sampling frame. MFI staff may have reasons for excluding groups that have the effect of biasing the survey results. Within an area, interviewers may purposely skip less accessible client households or may tend to select better-built dwellings while discounting the poorest accommodations of Nonclient households. All of these practices can be major sources of error.
- *Nonresponse errors.* Nonresponse errors usually occur when households are not at home or they refuse to participate. To avoid these errors, interviewers will need to revisit the household at a later time or, if the household has refused, select the next household on a reserve list for clients, or select the next household directly following the Nonclient household.
- *Interviewing errors.* Interviewers who conduct interviews in an awkward, tiring, or offensive manner can jeopardize the quality and extent of cooperation of the respondent. To avoid these errors, interviewers should know all questions thoroughly, so that they ask the questions exactly as worded, and should also know the sequencing of the questions and maintain this order at all times. Interviewers also can improve responses by helping respondents to understand the questions. These probing techniques help to motivate the respondent and also focus him or her on the specific information being asked. Finally, if any changes need to be made to the questionnaire, interviewers need to keep track of these; and all interviewers then need to agree to enact the changes uniformly and simultaneously. A cap on the maximum number of questionnaires completed per day may be considered to ensure that questionnaires are not completed hastily.

4.5.2 Stage two: Understand the content of the questionnaire

Format

Before this training, all local adaptations to the questionnaire should be drafted by project staff and the edited version of the questionnaire used for training. In training staff to use the questionnaire, make sure they understand its content and how responses should be recorded:

- Review all questions to identify their purpose,
- Clarify any definitions involved as they have been adapted for local conditions,
- Differentiate between the choices of responses,

- Describe how to record answers using codes,
- Explain the ordering of questions.

In general, this stage of training is best kept informal with plenty of opportunity for questions. Any confusion or skepticism on the part of interviewers over the wording of questions, the nature of the response groups, or the flow in the questionnaire can signal problems in it. Encourage evaluators to be forthcoming about how improvements can be made.

Materials

Training materials for this stage are a copy of the revised questionnaire for each participant and extra paper for trainees and trainers to take notes.

Time

This stage of training will take a half-day.

4.5.3 Stage three: Standardize translation of the questionnaire into local language or languages

Format: Small groups to translate, large group discussion to review translations
Materials: Extra copies of the questionnaire, sheets to make notes
Time: Half day

Many MFIs operate in areas where more than one language is spoken. In most cases, respondents will be better able to understand the questions if their local language is used in the survey interview. Successful implementation of a field survey in more than one language requires working through the wording for each question in each language with all interviewers to make certain that all responses are consistently measuring the same thing. Translation may not necessarily require that the translated version be a written one. If interviewers are highly skilled and truly bilingual, and can read English proficiently enough to translate accurately at the time of the interview, a standardized verbal translation may be adequate.

The format for this stage will involve first breaking interviewers into small groups according to their language skills. If the local language skills of the interviewers are not found to be adequate, it may be necessary to use outside expertise for language translation. Each small group will review the appropriate wording in each local language to be used for each question and response category. Notes of the words chosen can be recorded on extra sheets of paper. Once a draft translation is compete, a round table review of each translated version by all survey personnel can check for consistency.

Once translations are agreed upon, key words to each question can be listed on a separate sheet of paper and their translation into other languages shown. This is one way to avoid having to translate the entire document. These sheets can be used as references during actual interviews.

4.5.4 Stage four: Practice interviewing in local language or languages

Format: Small groups of three interviewers to rotate roles of interviewer, respondent and observer
Materials: Copies of the questionnaire, translation notes
Time: Half day

Good training programs provide extensive opportunity to practice interviewing. Practice helps interviewers to:

- monitor for consistency across languages,
- build familiarity with the exact wording and flow of questions,
- practice complete and accurate coding of questionnaires, and
- build confidence in their interviewing skills.

The format for this stage of training can be varied. Role-playing by pairs of interviewer and respondent is effective, both with and without an observer. Switching roles and partners can add variety to the needed repetitions of practice. All participants should be strongly encouraged to provide constructive feedback to their colleagues on how to improve their skills.

4.5.5 Stage five: Pretest the questionnaire

Format: Field travel to nonselected site, interviews conducted in groups with two interviewers and one observer. Follow-up discussion of any problem areas and needed changes
Materials: At least three questionnaires for each enumerator
Time: Full day

Pretesting a questionnaire in the field is a standard practice for finding weak points in the questions being asked, errors in the logistical plan, and additional need for field staff training. It provides an opportunity to make corrections before doing the actual survey. Pretesting usually involves all levels of the survey team. For questionnaire designers, it is a chance to see if the wording of questions is appropriate and interpreted consistently. Pretesting of the questionnaire will uncover at least a few unforeseen responses that do not fit into existing response categories. These can either be used to add to the list of existing codes, or to revise the definition of one or more codes already set. The pretest also cues project managers and field supervisors on the time commitment and resources required to locate and interview respondents.

The pretest is also an important training tool for interviewers who practice sampling methods and gain confidence in conducting the interviews. Each interviewer should have enough time to conduct at least two supervised interviews, preferably three if time permits. Supervisors monitor the interviews and completed questionnaires from each interviewer and give feedback on how he or she can improve performance. In addition to individual feedback, in a group format, interviewers will benefit from an opportunity to share their experiences, ask questions, and make suggestions for improvements. Interviewers will likely have questions related to the random walk method of

sampling Nonclient households. These need to be addressed and the rules for implementing the technique reviewed.

Pretesting the questionnaire involves more than finding households to ask questions. A pretest survey site should be selected where nonsampled MFI clients are located. The pretest should include an opportunity for field teams to practice random sampling at the individual client and Nonclient level. It should include a visit to local leaders to experience their reaction to the proposed interviews.

Day 1	Day 2 &3	Day 4
Review study background -- Purpose of the study, -- Sampling frame to be used for identifying households, -- Field operations plan, -- Role of the interviewer and principles of good interviewing, -- Potential sources of error.	Review content of the questionnaire Standardize translations of the questionnaire in local languages Practice interviewing in local language.	Pretest the questionnaire Gather feedback on needed changes.

PART II

ANALYZING THE DATA

CHAPTER FIVE:
MANAGING THE SURVEY DATA

Transforming written information from a questionnaire into a structured electronic database involves several meticulously executed stages of work. First, electronic variables must be defined and data types determined. Second, the data themselves must be entered into spreadsheet files, with each file clearly linked to all others. Once data are entered, they must be cleaned of errors so that the results of data analysis can be regarded as accurate.

Successful data management requires specialized skills in computerized spreadsheet software. This manual provides (on the accompanying diskette) data-entry file templates that can be adapted to fit the customized questionnaire in any given country, but the person carrying out these adjustments should have an understanding of how the data eventually will be used.[3] We recommend that an experienced data analyst be tasked with adapting the data-entry files and that this same person supervise individuals entering the data. Data entry does not require specialized computer skills, but individuals for this job should have some computer background in managing files on a personal computer and good typing skills. Poor typing usually translates into more time needed to complete the process and more data-entry errors. Data cleaning incorporates techniques best handled by an experienced data analyst with a background in statistics. This person can be assisted by data-entry persons to make the actual corrections.

The following sections describe in detail how data for this survey are entered into well-defined files and then cleaned of errors. This manual assumes that data analysis will require use of SPSS software and therefore provides guidance in the data-entry and cleaning processes for this software.

5.1 Data file structures and database design

5.1.1 Structuring files

Entering all questionnaire data into the same spreadsheet file would make the eventual analysis of data inefficient and disorderly. To avoid this, several different spreadsheet files can be defined and specific variables entered into each. Four separate files are needed to manage the data contained in the questionnaire:

- *F1: Household-level data file.* This file records all data collected at the household level (all sentence-type questions with only one response per household—all responses except those recorded in Tables B1, B2 and E2).
- *F2: Adult data file.* This file records all data collected at the adult member level (Table B1 in the questionnaire).

[3] File information on the templates are provided in Annex 5.

- *F3: Child data file.* This file records all data collected at the child member level (Table B2 in the questionnaire).
- *F4: Asset value data file.* This file records all data collected at the individual-asset level (Table E2 in the questionnaire).

Data-entry shells, or templates, for these four files are provided with this manual on diskette. These files can either be edited to reflect changes to the questionnaire or used as guides to create new data files from scratch.

5.1.2 Linking files within a relational database

Successful data management requires that all data pertaining to a particular household respondent be recorded under that household's identity code. This is achieved by creating spreadsheet files where each *row* in a spreadsheet is treated as "case" and contains information only for one household. The row always begins with the unique identity code assigned to that household.

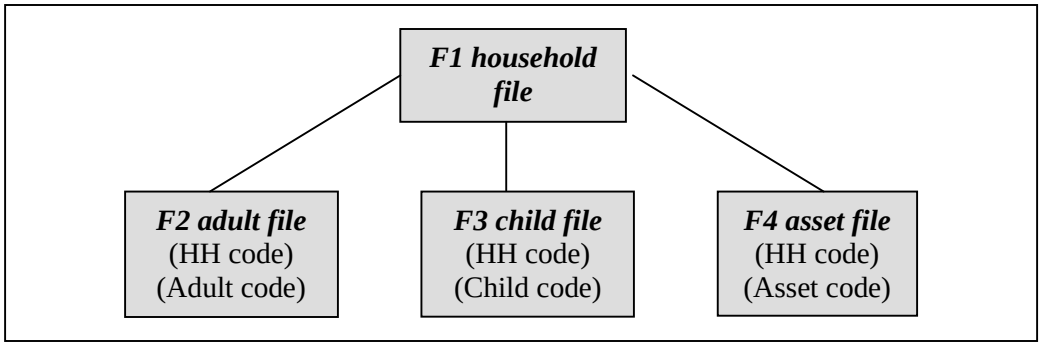
Household data can be recorded into separate files if more than one row in each spreadsheet contains information on the same household. This situation exists for questions pertaining to household members and assets. Each case within a spreadsheet file must be uniquely identified by replicating in each file the household identity code for each case, and adding an additional identity code to further identify all additional data specific to a single household. For example, if a household contains more than one adult, then the household code and a unique code for each adult are used to record information about that adult.

Linking files through overlapping case identity codes is essential to support analysis of the data. The researcher will need to pool information contained in individual files to conduct core analysis. The unique identity codes provide the means of linking these files together. (Additional codes, such as those for the survey area, will also help to categorize households; however, these will not be unique to each case.) The unique identity codes for the four files are as follows:

- F1: household identification code
- F2: household identification code + adult identification code
- F3: household identification code + child identification code
- F4: household identification code + asset identification code.

Figure 5.1 summarizes the data file structure proposed for this study.

FIGURE 5.1 Relational file structure within SPSS database



5.2 The general organization of the SPSS program

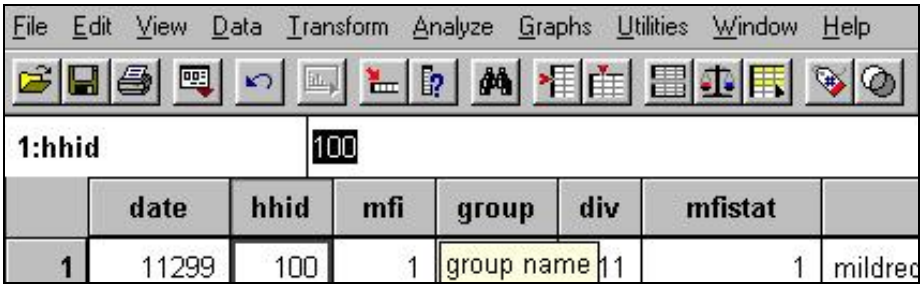
The SPSS program is Windows based and can be operated largely from its menu system, complete with toolbars and icons.

5.2.1 Main menu bar

The SPSS main menu bar is displayed across the top of the opening window when the user first starts the program. The main menu bar (Figure 5.2) has the following menus:

- **File**—used to create new files, open and save existing files, and print
- **Edit**—used to undo, cut, copy, paste, and clear; also used to find data
- **View**—supports customization of SPSS's appearance on the computer
- **Data**—accesses commands to define and sort variables, merge and aggregate files, and select and weigh cases
- **Transform**—accesses commands to transform or convert variables to another form
- **Analyze**—accesses different options for statistical analysis techniques
- **Graphs**—used to create bar, line, area and other graph types
- **Utilities**—used to find information about variables and files
- **Window**—used to switch from one window to another
- **Help**—offers tutorials, help by topic, a syntax guide, and search by help topic
-

FIGURE 5.2 SPSS main menu



We strongly recommend that each SPSS user go through the complete online tutorial provided with the SPSS software before proceeding with this manual. The rest of the SPSS information covered in this manual focuses on specific applications used to support the MFI poverty-assessment tool.

Data can be entered into the computer through any of a wide range of software packages. The most common means of entering smaller data sets, such as the one for this survey, is a spreadsheet program, such as Microsoft Excel; or relational database program, such as Microsoft Access. Many of the instructions and clarifications provided in the following section can be applied for use with

these or other spreadsheet software programs. Regardless of the data entry program used, however, analysis will eventually require SPSS or SAS, two statistical packages that support principle component analysis.

5.2.2 The views of SPSS

In addition to its main menu system, SPSS also has three windows for displaying information about data. The DATA VIEW window displays the actual data in spreadsheet form, as shown in Figure 5.2. The VARIABLE VIEW window (Figure 5.3) shows information on variable definitions. The variable view of the data file is used to add and delete variables or change the characteristics of variables. In this view, each row summarizes information about a single variable and each column lists a characteristic of that variable. In both data and variable views, users can add, change, and delete information contained in the file.

FIGURE 5.3 SPSS VARIABLE VIEW window (version 10 only)

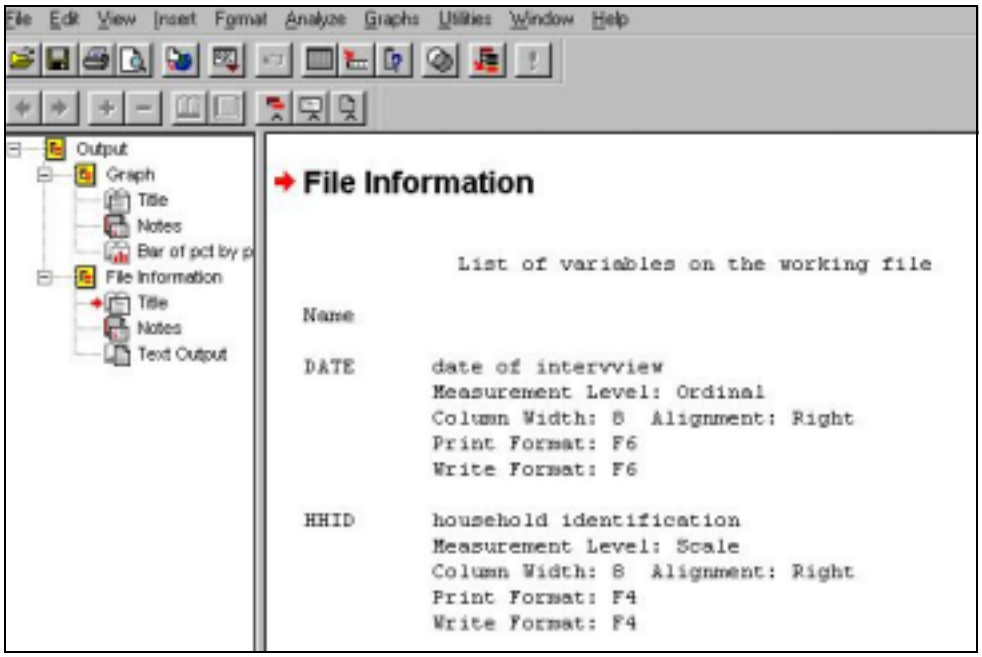
File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	date	Numeric	6	0	date of interview	None	None	8	Right	Ordinal
2	hhid	Numeric	4	0	household identificati	None	None	8	Right	Scale
3	mfi	Numeric	2	0	MFI survey area code	{1, North}...	None	8	Right	Ordinal
4	group	Numeric	3	0	group name	{11, United wo	None	8	Right	Ordinal
5	div	Numeric	3	0	community name	{11, River F ...	None	8	Right	Ordinal
6	mficlien	Numeric	1	0	MFI client status	{0, non-client o	None	8	Right	Ordinal
7	hhldrepl	Numeric	1	0	whether household iis	{0, no}...	None	8	Right	Ordinal
8	orighhld	Numeric	1	0	what the original hous	{1, not found}..	None	8	Right	Ordinal
9	memnths	Numeric	1	0	months of membershi	None	None	8	Right	Ordinal
10	respo	String	20	0	name of respondent	None	None	20	Left	Nominal
11	hhead	String	20	0	name of hhold head	None	None	20	Left	Nominal
12	hhaddres	Numeric	5	0	household address	None	None	8	Right	Scale

The third type of window is the OUTPUT VIEW window (Figure 5.4), which actually displays the contents of a separate file. As procedures are run in SPSS, the results are automatically displayed in the output view file. In output view, the window is divided into two parts. The left section contains an outline of the output contained in the file, and the right section contains the tables and charts created by the user and can be used to locate and move to different output contained in the file. In Figure 5.4, the small arrow shown left of the word 'title' corresponds to the title displayed in the right section and marked by a large arrow. Users can use the scroll bar to browse the results or double click on the icon to move to a particular output.

As new procedures are run, the resulting tables and graphs are added sequentially to the output view file. The output file can be saved and reopened by assigning a name and the extension .spo (which is automatically added any time files are saved). The tables within SPSS output files are transferable to most word-processing programs and can easily be added directly into a summary report.

FIGURE 5.4 SPSS OUTPUT VIEW window



5.3 Data entry methods for survey data

5.3.1. Preparation of data-entry forms and file documentation

To enter data, a variable name is needed for each possible type of data collected on the questionnaire. Each *column* in the spreadsheet is labeled with the name of one of these variables; and these variable names, or labels, usually are sequenced according to the order in which they appear on the questionnaire. An electronic data entry template for each of the four spreadsheet files accompanies this manual (see annex 5 for file information). Each file can be edited to reflect local questionnaire adaptations.

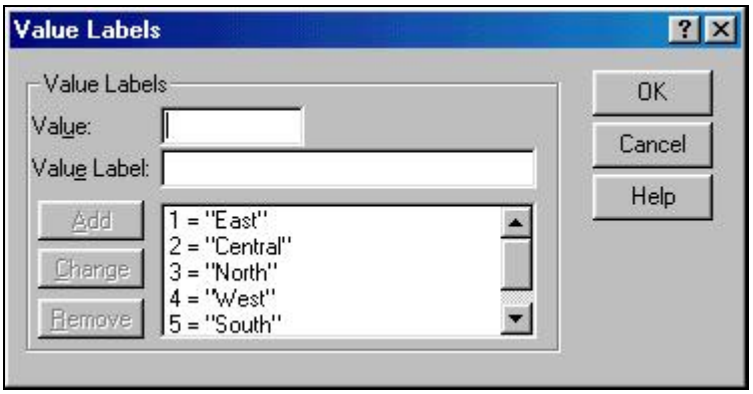
Most spreadsheet and statistical software programs automate the variable-creation process so that the definition of the variable can be entered at the same time that the variable label is created. The variable definition not only specifies the meaning of the variable label but also designates the format of the variable data, and lists code numbers and what each code represents for all pre-coded responses.

Variable type refers to whether data are numbers, dates, currencies, or strings. Most variable types in this questionnaire are numeric so that they can be analyzed using statistical procedures. All data containing letters or a mixture of letters and numbers are categorized as strings.

Variable and value labels are created in the VARIABLE VIEW window. A variable name must begin with a letter and cannot exceed eight characters. In addition to a variable name, a variable definition and value labels representing different response codes also need to be specified in the DEFINE VALUE LABELS dialogue box (Figure 5.5). It is accessed from the VARIABLE VIEW window by clicking on the right end of the values cell for that variable.

Variable measurement records the way in which data are measured for each variable. SPSS defines data as nominal, scale, or ordinal. The terms are shown in the far right column of Figure 5.3.

FIGURE 5.5 Dialogue box for SPSS value label procedure



Variables are analyzed according to the type of measurement used. Variables created from the questionnaire constitute three types of measurement:

- *Nominal data*, where the number codes recorded represent labels for categories of responses and tend to identify and classify information. Examples from the questionnaire include marital status, gender, and location codes. The code numbers do not represent systematic sequencing that reflect an underlying scale.
- *Ordinal data*, where the sequence of numbers for a variable reflects an ordered relationship. Code responses for ordinal data are assumed to measure points along an underlying continuous function that may specify graduations in quality or cost. Examples of ordinal data are education levels of adults, or the quality of drinking water and toilet facilities.
- *Interval- and ratio-scaled data* are like ordinal-scaled data except that the differences between numbered values are equal. Ratio data is a specific type of interval data, but requires an absolute zero point. These data represent an actual unit of measurement, such as quantity, size, weight, or distance. Expenditures on clothing and footwear is an example of a scale variable.

Whenever possible, quantifiable variables are recorded as interval or ratio data, since this type of data permits a more rigorous statistical testing.

A list of all variable definitions and value labels can be summarized in SPSS by selecting the **Utilities** menu and the **File Info** option. Figure 5.4 shows an example of this list. The printout of this list shows all variable names, their definitions, types of measurement and all value labels for

nominal and ordinal data. The list should be carefully proofed and edited before any data is entered. Data entry cannot begin until the data-entry files exactly mirror the contents of the adapted questionnaire. This will mean that all variable names and definitions, and all code numbering and labels correspond to adaptations made to the questionnaire.

5.3.3 Entering the data

Before actual data-entry can begin, all data-entry personnel should be trained in using the data-entry files by practicing entering actual questionnaire data. During this trial stage, each data-entry person has the opportunity to practice how to edit, save, and reopen the data files. Data can be entered twice to check for consistency and accuracy.

Data entry is best achieved by following systematic procedures to minimize data-entry errors. In general, the data-entry person should enter all data from a questionnaire before moving to the next questionnaire. An exception to this rule can be made if the data entry is to be organized by file rather than by questionnaire. In both cases, the data-entry person should make notes on each questionnaire that show who entered the data and which part of the questionnaire was entered.

Missing values on questionnaires require special procedures for accurate recording. Missing values can be of two types. First, in some cases, no answer is required from some households and the interviewer has purposely left the answer blank. This is not a true missing value, and in the data-entry file, the cell for this variable and case should be left blank. In SPSS, a decimal point will appear in the middle of the cell to mark it as empty. The computer will treat this cell as a 'sysmis' or system missing.

In other cases, a value is called for but none provided. This is what can be considered a missing value and efforts should be made to determine the correct value that belongs in the cell. If none can be found, the cell is also left blank and treated as a 'sysmis'. Under *no* circumstance should the number "0" be used to record a missing value, as "0" constitutes a reasonable response for many quantitative variables. Ideally, missing values should also be referred back to the interviewer to see if he or she can recall the correct responses.

5.3.4 Making electronic backups

Entering data is a time-consuming process. Once entered, data will be cleaned and further prepared for analysis, all of which takes time. Successful data management entails backing up all files regularly and storing copies in several different locations for safekeeping. Once data are entered and cleaned, master copies of these data files should also be saved and safeguarded. Subsequent alterations to these original versions should be saved and safeguarded under different file names.

5.4 Cleaning the data

An effective means of detecting errors with minimum effort is to enter all data twice, each time into a different file. Cases can then be compared between the two files to see where differences occur.

Because data entry is estimated to take no more than three to four days, entering twice can be more cost-effective than searching for errors at later stages.

Data-entry errors can occur in several ways. First, codes may be entered for a variable that do not exist in the original questionnaire. Second, the person entering the data may key in values for a variable incorrectly. Data errors can also occur in the questionnaire at the time of the interview that were not caught in the field and not detected by the person entering data. These can include values that are inappropriate for the question, or values that contradict information captured by other variables. Finally, data may be missing for variables in some questionnaires either because the respondent failed to answer the question or the interviewer failed to record the answer.

5.4.1 Data cleaning procedures

Data cleaning consists of a series of procedures that locate the various types of errors described above and guide the cleaners in how to make corrections where appropriate. These procedures are briefly outlined below.

Wild codes. We want to rid the data set of all codes that do not exist for a particular variable. One method for identifying such "wild codes" is to test for frequencies on each indicator and compare the value codes for each to what was written in the original questionnaire. When wild codes are found and data assigned to these, then the original questionnaire must be used to reenter the correct code value. The SPSS procedure for frequency testing is described in section 5.5.

Consistency checks. Checks on the logical patterns of answers can also be used to find data errors. Consistency checks can be done in several ways within SPSS. One method is first to filter the data set for only cases responding in a certain way, and then running a frequency test on a second variable to check for inconsistencies.

In our example, households indicating they had no food shortages in the past year (C8) would not indicate they did not have enough to eat in the past month (C7). If the data are clean, all households answering zero to C8 will also show a zero response to C7. A frequency test of responses to C7on all households answering zero to C8 would uncover any inconsistent responses.

Another method for checking inconsistencies across variables with only a few categories of responses is to run cross tabulations where the responses for one variable are cross-checked in tabular form against the responses for another variable. An example here would be to check that households who cook with electricity (D8) also report having access to an electricity supply (D7). The SPSS routine for running a frequency test is discussed in the next section. Procedures for running a cross tabulation are described in section 7.2.2.

Extreme case check. In some cases, responses to a question can seem highly improbable either because they are extreme when compared with the responses given by other households, or because they seem improbable given other responses from the same household. In a survey, a household with few assets, limited food supplies, a poor diet, and low expenditures on food and clothing was found to hold land assets worth nearly $500,000. Not only was the value much higher than all other households in the survey, but it also seemed inconsistent with the household's other responses. It was found that the data-entry person had typed too many zeros in the landholding-variable cell.

Extreme cases can be identified through several techniques in SPSS. Perhaps the easiest is creating a "box plot" of variable responses, as described in section 5.5.4.

5.4.2 Correcting data errors

Procedures for correcting data depend on the source of data error. In most cases, the source cannot be determined without checking the actual questionnaire. If the response written on the questionnaire is different than the number entered in the file, then the error is from data entry. The error is corrected by changing the number in the file to that shown on the questionnaire.

If the response shown on the questionnaire is the same as entered in the file, two scenarios are possible. First, the number may not be an error but simply an unusual response. To verify this, look through the household's responses to other questions to see if the response is plausible for that household. If the response seems to make sense, then leave the response unchanged. The second possibility is that the response seems unreasonable, in which case the cell is emptied of the false response and treated as a sysmis, or system missing response. A decimal point will appear in the empty cell.

5.5 Applying SPSS procedures for cleaning data

Three common SPSS procedures for cleaning data are frequencies, descriptives, and box plots. In addition, selecting subsets of cases is required to locate the case or cases with data errors.

5.5.1 Locating cases with data errors

Selecting subsets of data files is a useful technique for data cleaning. Generally it is used to restrict analysis to only a specific group of cases. In data cleaning, it is used to locate the case or cases containing errors.

Selecting subsets of cases is easily done using SPSS menus. Begin by selecting from the main **Data** menu the option **Select Cases**. This opens the SELECT CASES dialog box. The default for selecting cases is set to include all cases ("Select: All cases"). To change the default, click on "If condition is satisfied," as is shown in Figure 5.6, making the "If" button available for selection. Click on the button to open the SELECT CASES: IF dialogue box.

Figure 5.7 shows the SELECT CASES: IF dialogue box. Scroll down to find and highlight variables from the list provided at the left of the form. Click on the arrow button to move variable names into the box to the right, and select from the operator and number keys the components needed to build an equation setting the rule for selecting cases. Common operators used in rules are >, <, =, and ~= (not equal). For example, "mfi = 1" selects the cases in the first MFI survey area. Click on "Continue" to return to the previous menu and check that "Filtered," under "Unselected Cases Are" is selected (the circle next to it is filled in). Click on "OK" to run the selection command. Note that clicking on "Deleted" under "Unselected Cases Are" removes from the data file all cases that have been deselected. Do not select the deleted option unless you have an unusual reason for doing so.

In our example, once the number of cases is limited to mfi cluster 1, use the frequency procedure to check that no localities or group codes appear outside the range assigned for that area.

FIGURE 5.6 SPSS Select Cases dialogue box

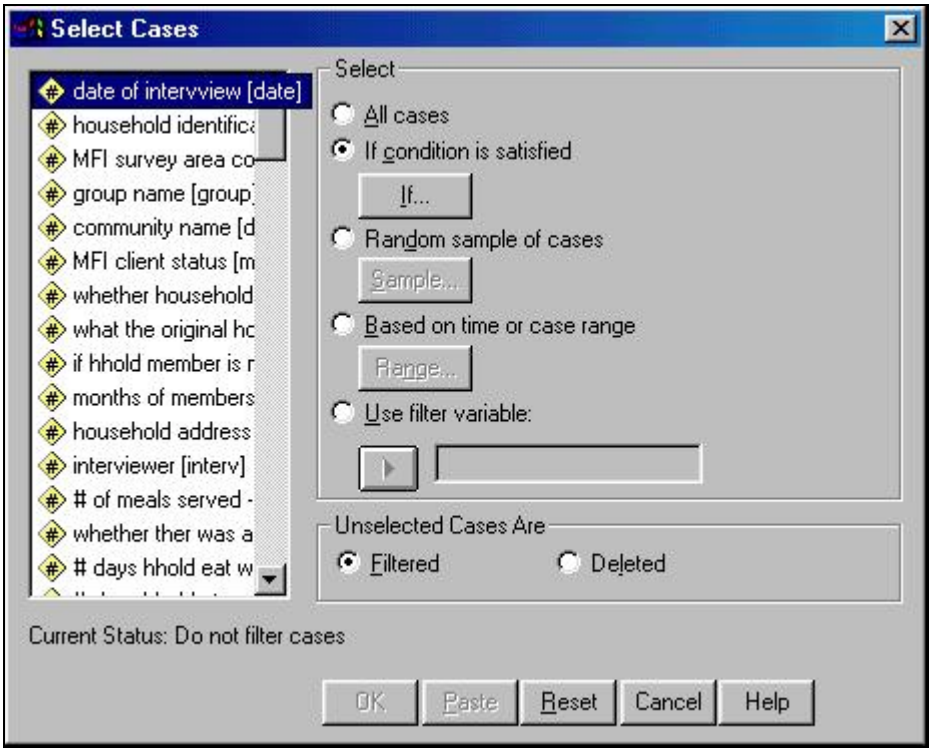
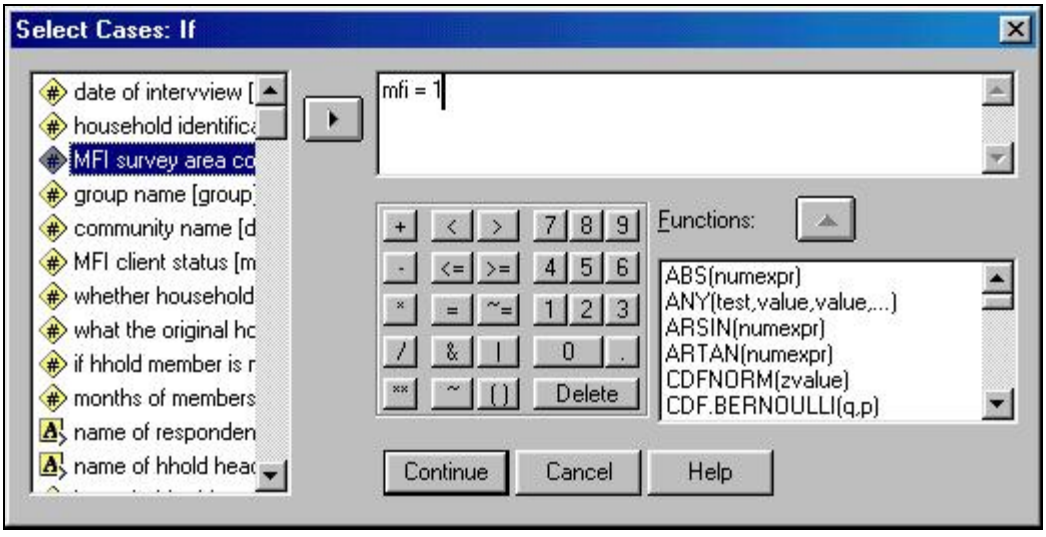


FIGURE 5.7 SPSS Select Cases: If dialogue box

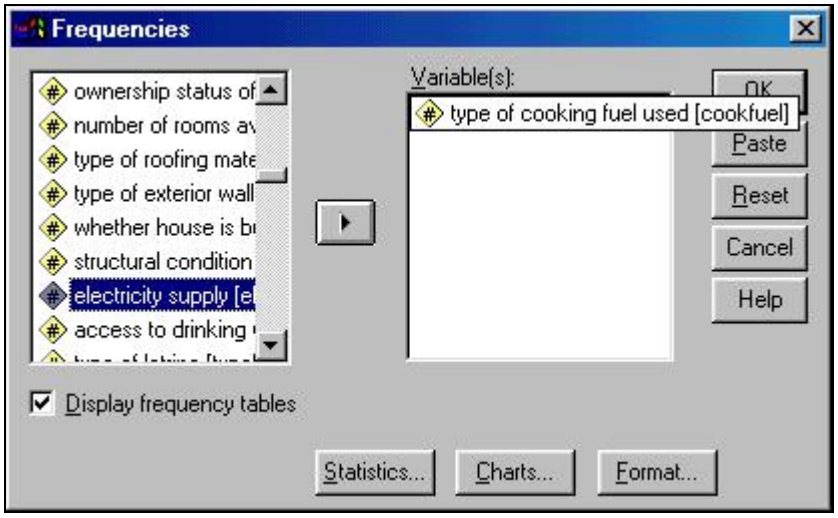


SPSS returns to the SELECT CASES: IF dialog box, which displays at the far left of the data file a diagonal slash through the case number of deselected cases. To deselect all cases, return to the SELECT CASES dialog box and click on "All Cases," then "Continue."

5.5.2 Frequencies

The frequency function in SPSS can be used to determine the frequencies of different responses for a variable. To compute frequencies in SPSS, click on **Analyze** in the main menu, then **Descriptive Statistics**, then **Frequencies**. This will open the FREQUENCIES dialogue box (Figure 5.8). Click on the variables to analyze from the list to the left, then click on the arrow key to move them to the "Variable(s)" box at the right.

FIGURE 5.8 SPSS dialogue box for testing frequencies



The frequency procedure can be further specified to present results as a chart (an example is shown in Figure 5.9) or to produce specific statistical summaries. The bottom of the window contains the following buttons that can be used to refine the frequency procedure:

- "Statistics"—to access a dialog box to select types of descriptive statistics
- "Charts"—to access a dialogue box to chart frequency distributions
- "Format"—to access a dialog box to set the presentation format of results.

Table 5.1 shows the results of using the SPSS frequency function. Instead of variable names and value codes appearing in a table, SPSS displays the defined variable label and value labels. If these do not appear, they may not have been entered. From the variable view, click on the variable LABEL to add a variable definition dialog box and then click on "Labels." Click on the VALUES column to add value labels.

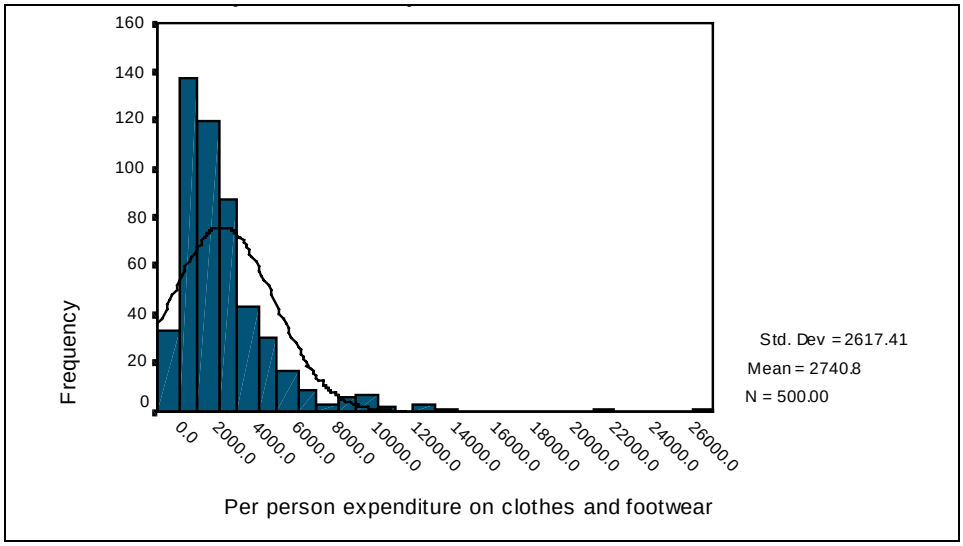
In Table 5.1, one case in the cooking fuel variable is missing a value. To find out which case contains a missing value, use the SELECT CASES: IF dialog box (accessed from the SELECT CASES dialog box) and set the "if" condition to "MISSING [cookfuel]." Run the frequency test for the household identification code to locate case codes have missing values. Then use the **Go To Case** option in the **Data** menu to locate each case. Once the case is reviewed, locate the original questionnaire to see if the correct response is provided. Table 5.1 shows that one case has a missing value for type of cooking fuel.

TABLE 5.1 Frequencies: Type of cooking fuel used by households

Response category	Frequency	Percent	Valid percent	Cumulative percent
Collected wood	179	35.8	35.9	35.9
Purchased wood or saw dust	128	25.6	25.7	61.5
Charcoal	108	21.6	21.6	83.2
Kerosene	61	12.2	12.2	95.4
Gas	15	3.0	3.0	98.4
Electricity	8	1.6	1.6	100.0
Total	499	99.8	100.0	
Missing system	1	0.2		
Total	**500**	**100.0**		

Frequency charts are useful to visually inspect the distribution of responses for a single variable, which can help in finding outliers within the data. Figure 5.9 charts the distribution of per-person expenditures on clothing and footwear, and also lists the mean and standard deviation for the variable. Because the variable has a large number of different values, the data were graphed in range segments. The chart shows that distribution is slightly skewed to the right, with two responses much larger than all others. These may or may not represent data errors. To find out, isolate the households recording expenditures per capita greater than 2000 and then check if other responses from the households indicate higher levels of wealth.

FIGURE 5.9 SPSS graph of distribution of responses by amount of expenditure



5.5.3 Descriptives

The descriptive function calculates almost all the statistics provided by the frequency function but provides a compact table of statistics. Descriptive tables are made by clicking on **Descriptive Statistics** in the **Analyze** menu, then **Descriptives**. Table 5.2 is an example of a descriptive table that shows an unusual outcome: that a household could spend nothing on clothing and footwear for entire year. Inspection of the questionnaire and discussions with the field supervisor determined that the value was not erroneous, simply unusual.

TABLE 5.2 Descriptive statistics for per person expenditures on clothes

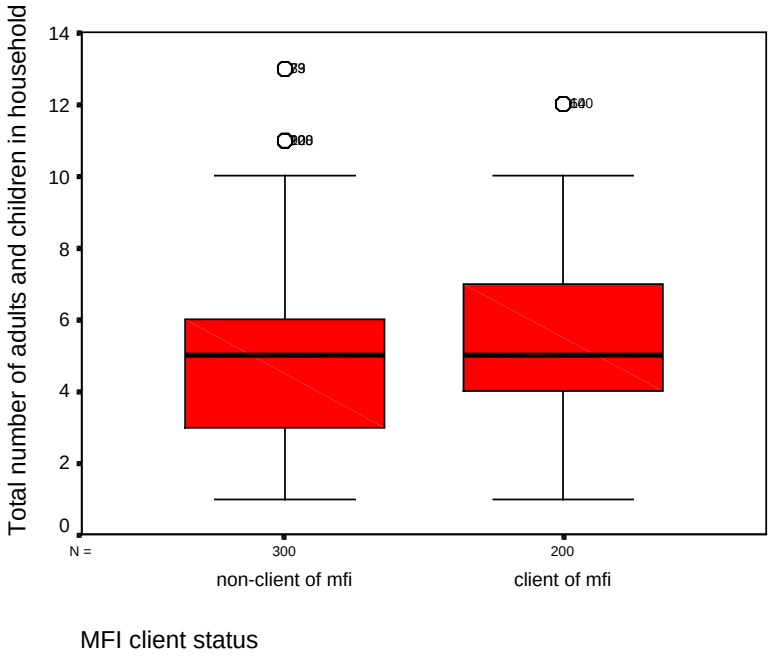
	N	Minimum	Maximum	Mean	Std. deviation
Per person expenditure on clothes and footwear	500	0.00	26666.67	2740.8159	2617.4092
Valid N (listwise)	500				

5.5.4 Box plots

In addition to frequency and descriptive tables, data can be explored through the use of a box plot of the data. Instead of plotting actual values, the box plot shows the median, 25th percentile, the 75th percentile, and values that are far removed from the rest. In Figure 5.10, the thick line near the middle of each box represents the median for each group of cases. The box area represents the range in which 50 percent of all cases in each group fall. The box plot includes two categories of cases with outlying values. Cases with values more than 3 box-lengths from the upper or lower edge of the box are called "extreme values," and are marked with an asterisk (*). Cases of values between 1.5 and 3 box-lengths from either edge of the box are called "outliers" and are marked with the letter "O."

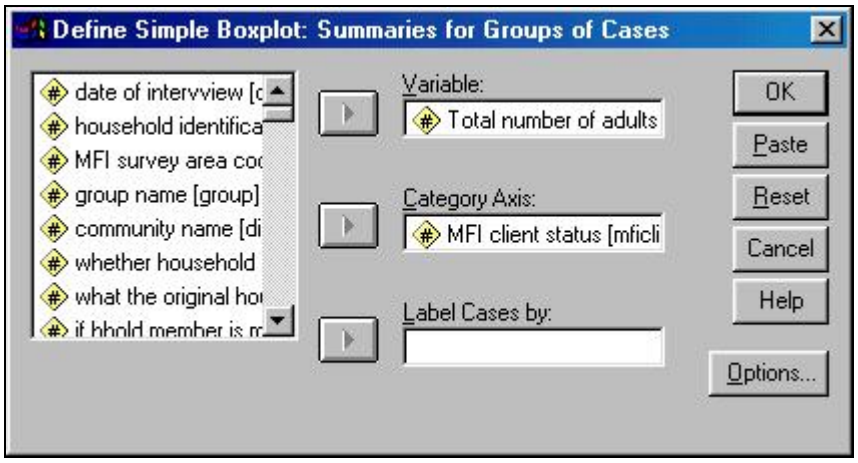
Box plots can be used to clean data by identifying extreme outliers. Box plots are also useful for comparing the distribution of values among several groups. The box plot in Figure 5.10 graphs the number of adults and children in each house by grouping data according to client status. As the graph shows, for both clients and nonclients the median family size is roughly five, with only three outliers and no extreme values detected. These results would not indicate a data-error problem.

FIGURE 5.10 Box plot of case data collected on household family size



Box plots and other graphs are created by selecting **Boxplot** from the **Graphs** menu. Choose the "Simple" as the graph style and "Summary for groups of cases." This will open the dialog box shown in figure 5.11. Choose the variable to be plotted and the groupings to distinguish cases.

FIGURE 5.11 SPSS dialogue box for creating box plots



5.6 Data-cleaning routines

Once errors are located, they are corrected in the original data files. This section outlines specific tasks to follow when cleaning data for each of the four types of data files. The list is not exhaustive. These tasks are presented as examples of how data can be cleaned. Additional checks are certainly needed that apply the concepts described below to test for errors in different variables.

5.6.1 Household file (F1)

Conduct frequency tests on all variables with a limited number of response values. These include all variables with categories of responses defined. Only variables measuring an actual number need scrutiny (used for questions C2, C3, C5, C6, C10, C11, and D2 in the questionnaire).

Verify that no household has answered for both questions C2 and C3. You can do this by selecting cases where the variable for C2 is "1," and creating a frequency table for responses to C3 to see that no cases are listed. (All cases should have a missing value.)

Verify that no household has reported an unusually low or high number of meals in the past two days (questions C2 and C3). Use a frequency table or box plot to check for outliers. When outliers are found, check with the original questionnaire to determine if they are unusual or erroneous. Change the value if a data entry error is found or delete the value if no correction can be made.

For questions C5 through C6, verify, by using a frequency table, that no answer exceeds "7," the highest number of days possible, and no answer is less than 0. For question C10, verify with a box plot that no unusual outlier responses have been recorded for weeks of local staples stored. For question C11, verify that no response exceeds the maximum amount of money stated in the question.

Finally, for question D2, verify with a box plot that no unusual outlier responses occur for the number of rooms in the household's dwelling.

5.6.2 Adult file (F2)

The adult file repeats the client-status variable from A6. Verify that the households listed as MFI clients are actually clients of the MFI. This can be done by checking that all households listed as MFI clients also report at least one member as a client in table B1. The reverse procedure can be used to check that households listed as nonclients are actually nonclients. This can be done by selecting only cases of nonclients and creating a frequency table for adults who are members.

Create frequency tables on all variables with a limited number of response values. These include all variables *except* age and expenditures on clothing. Create descriptive tables or box plots to check for outliers for these two variables. Any cases where ages are recorded as under 15 would indicate an error. Ages over 100 would also be questionable. Clothing expenditures well above the range of most households should also be checked to verify that these households also indicate a higher level of wealth from their responses on food consumption, housing, and ownership of assets.

Verify that each household head has been correctly identified. No individual with an ID code of "1" (for head of household) should have a response under the variable for "relation to head," and no household should have more than one member with an ID code of "1." Use select cases to filter only cases with an individual ID code of "1,"or head of household then create a frequency table using the variable relation to household head.

5.6.3 Child file (F3)

Create descriptive tables or box plots to check for outliers for the two variables for ages and expenditures on clothing and footwear. Any cases where ages are recorded as over 14 would indicate an error. If found, select cases where "age > 14," and run a frequency test for the household ID code. As before, check that any unusually high expenditure levels on children's clothing and footwear coincide with higher expenditures for other children and adults in the same household.

5.6.4 Asset file (F4)

Ownership of assets reported by households is likely to vary considerably among households. Errors occur, in part, from households distorting information on the number and value of their assets. Errors also occur from inclusion of assets that may not be completely owned by the household (for example, either purchased through credit or rented). Data-cleaning procedures for assets then would want to screen for unusual combinations of information. A household owning four vehicles worth an estimated value of $25,000 but not having any electronics or appliances might seem unreasonable. Use the "Select If" option to filter cases owning assets at unusually high values and use a frequencies test to identify household identification codes. Check each case for inconsistencies in responses to other questions.

CHAPTER SIX:
WORKING WITH DATA IN SPSS

Specific SPSS skills and techniques are required to prepare data for analysis once it has been cleaned. The data contained in four separate files are combined into a single file—an expanded version of the *FI* household file. This is achieved in SPSS by using first the procedure for aggregating data, followed by the procedure for merging files. Data recorded about adults, children, and assets are used to create new aggregated variables that record information at the household level. Once all data for the analysis is contained in a single file, several new variables are calculated from existing ones. The SPSS function for transforming data will be used for this task.

6.1 Methods for aggregating data to form new variables in SPSS

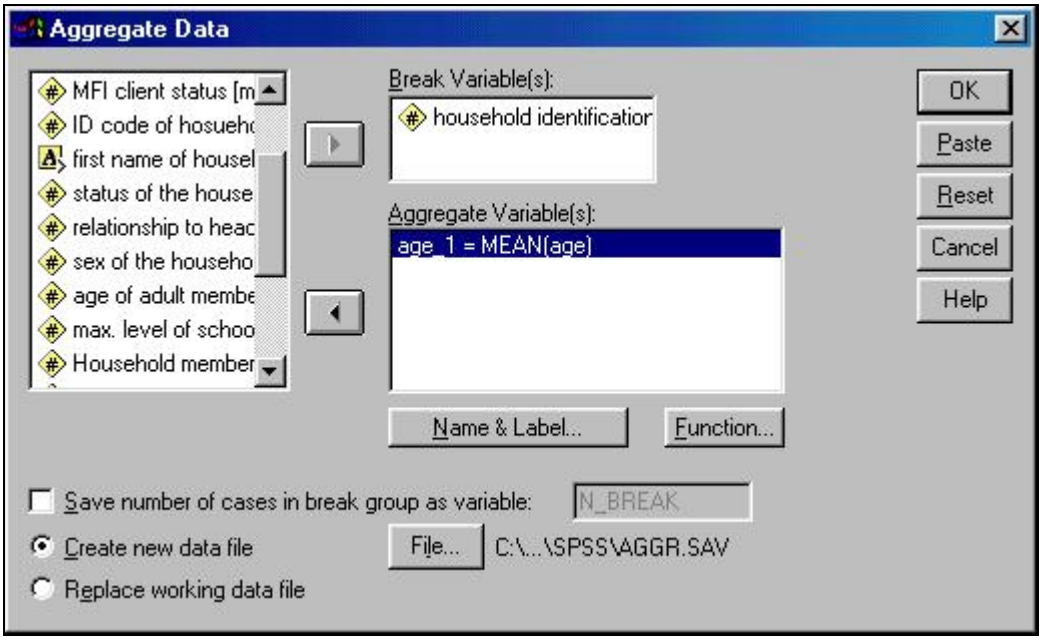
The data files for adult members of households (B1), child members of households (B2), and summary of individual assets (E2) all contain data to be aggregated at the household level. This process is needed to create household-level variables that can be analyzed with other variables already existing at the household level. For example, if the object is to know the number of adults in each household who can write, this information can be created from the adult file by counting the number of adults who answer yes to 'can write' in each household. However, the result is a number that is measured at the household level and therefore can no longer fit in the adult file. The aggregation function calculates the new variable, and the merge function transfers the new variable to the household file.

Aggregating data from the individual level or asset level to the household level requires that several steps be completed. First, the type of variable to be created from each is defined, and an SPSS function is set for calculating it. Second, the newly created variable is saved in a temporary file. Finally, the temporary file is merged with the household file by matching the household codes for each case.

6.1.1 SPSS aggregate data function

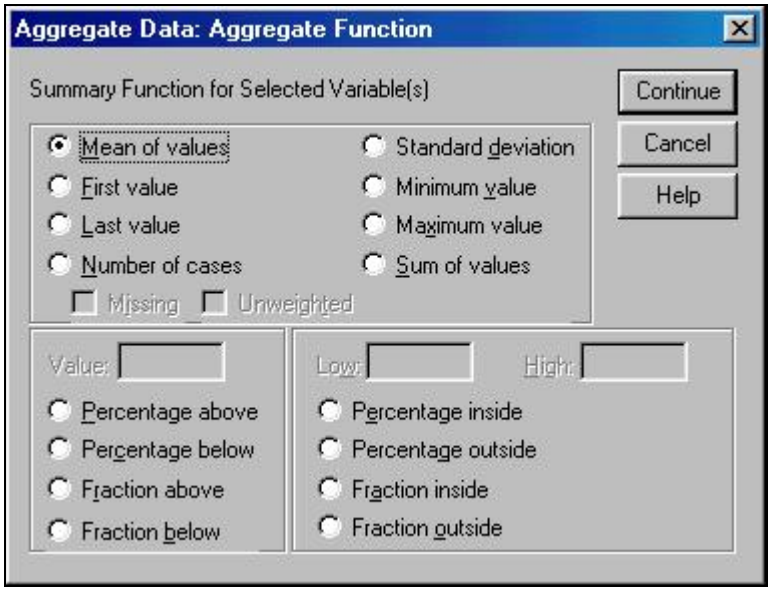
The aggregate process in SPSS requires that aggregate variables measured at the household level be temporarily saved in new files. The aggregate data function in SPSS is in the **Data** menu. In the AGGREGATE DATA dialog box (Figure 6.1), click on the variable name for household code in the list to the left and then click on the arrow key to move it to the box on the right labeled "Break Variable(s)." The "break variable" is the level at which data will be aggregated. All aggregations will use the household code as the break variable and all newly formed variables will be at the household level.

FIGURE 6.1 SPSS dialogue box for aggregating data



Move the cursor to the same variable list and highlight the variable to be aggregated. Use the lower arrow key to move this variable name to the box labeled "Aggregate Variable(s)." The "Name & Label" and "Function" buttons are now available. Click on "Function" to open the AGGREGATE DATA: AGGREGATE FUNCTION dialogue box (Figure 6.2).

FIGURE 6.2 SPSS dialogue box for selecting an aggregate function



In the AGGREGATE FUNCTION submenu (Figure 6.2), you can specify how to calculate each aggregated variable. Select "Sum of values" if a total number is needed for each household, select "Mean of values" if the average, or percent of total in each household is needed; and select "Number

of Cases" for variables where the number of occurrences within each household needs to be counted. In the example in Table 6.1, the function "Mean of values" was selected to calculate the average age of adults in the household.

6.1.2 Aggregating old variables to form new variables

Variables can be aggregated by a wide range of functions. The common methods used in this study record the value for each household case as one of the following:

- Mean, or the average of all individual or asset cases
- Sum, or the total of all individual or asset cases
- Count, or the number of occurrences of a particular response or condition.

TABLE 6.1 Aggregating from *F2*: Family structure for adults (aged 15 and above)

Individual variable	Output indicators	Method of aggregation
ID code of individual member	Number of adults per household (NUMADULT)	Select NUMBER OF CASES from the FUNCTION submenu to aggregate.
Age	Average age of adults (AGEADULT)	Select MEAN from the FUNCTION submenu to aggregate.
Maximum level of schooling	Number of adults with minimum specified level of education (EDUC1 EDUC2)	Select PERCENT ABOVE and specify a response code for education level as a cutoff point (completed high school or above, for example). Use different cutoff levels to create more than one aggregate variable.
Can write	Number of adults who write (NUMWRITE)	From the **Data** menu, select cases answering "yes";then select NUMBER OF CASES from the FUNCTION submenu to aggregate.
Main occupation	Number of adults with each main occupation (OCCUP1 OCCUP2)	From the **Data** menu, select cases for each response <7; then select NUMBER OF CASES from the FUNCTION submenu to aggregate. Result is creation of six indicators for each employment category.
Main occupation	Number of adults not working (NUMUNEMP)	From the **Data** menu, select cases in which the occupation code is > 6; then select NUMBER OF CASES from the FUNCTION submenu to aggregate.
Current client of MFI	Number of MFI members in household (NUMCLIENT)	From the **Data** menu, select cases answering "yes."Select NUMBER OF CASES from the FUNCTION submenu to aggregate.
Clothes/Footwear expenses	Total amount spent on clothes/footwear expenses (ADUEXPEN)	Select SUM from the FUNCTION submenu to aggregate.
Sex + head of household	Female-headed household (FHH)	From the **Data** menu, select cases where the relation to the head of the household = 1 (that is, head of the HH) and the sex is female. Select NUMBER OF CASES from the FUNCTION submenu to aggregate.

Individual-level indicators to be aggregated to the household level are listed in Tables 6.1 and 6.2. Asset-level indicators to be aggregated to the household level are listed in Table 6.3. It is easy to make errors in the aggregation process if the steps involved are not well thought through. The far left column lists the original variable used to create an aggregation. These are placed in the Aggregate Variables box. The middle column names and defines the output indicator that will be created for each household case. These will be saved in temporary output files. The far right column describes the procedures to follow in SPSS.

TABLE 6.2 Aggregating from *F3*: Family structure for children (aged 0 to 14)

Individual variable	Output variable	Method of aggregation
ID code of individual member	Number of children per household (NUMCHILD)	Select NUMBER OF CASES from the FUNCTION submenu to aggregate.
Age	Average age of children (AGECHILD)	Select MEAN from the FUNCTION submenu to aggregate.
Clothes/ Footwear expenses	Total amount spent on clothes/footwear expenses (KIDEXPEN)	Select SUM from the FUNCTION submenu to aggregate.

TABLE 6.3 Aggregating from *F4*: Value of selected household assets

Individual variable	Output variable	Method of aggregation
Value of individual animals by type (asset code = 1, 2, 3, 4)	Total value of livestock (VALANIML)	From the **Data** menu, select cases in which the asset code is < 5; then select SUM from the FUNCTION submenu to aggregate.
Value of individual transport assets (asset code = 5, 6, 7, 8, 9)	Total value of transportation assets (VALTRANS)	From the **Data** menu, select cases in which the asset code is > 4 and < 10; then select SUM from the FUNCTION submenu to aggregate.
Value of individual appliances and electronics	Total value of appliances and electronics (VALAPPLI)	From the **Data** menu, select cases in which the asset code is > 9; then select SUM from the FUNCTION submenu to aggregate.
Individual variable	**Output variable**	**Method of aggregation**
Value of televisions owned (asset code = 10)	Value of televisions owned (VALTVS)	From the **Data** menu, select cases in which the asset code is = 10; then select SUM from the FUNCTION submenu to aggregate.
Value of radios owned (asset code = 15)	Value of radios owned (VALRADIO)	From the **Data** menu, select cases in which the asset code is = 15; then select SUM from the FUNCTION submenu to aggregate.

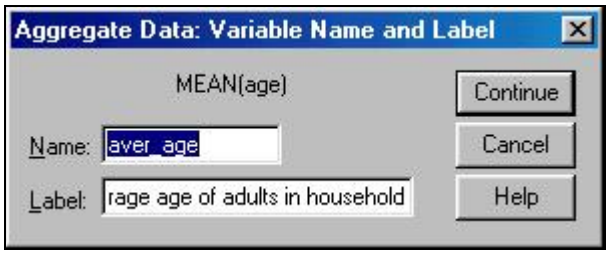
Aggregating data on assets will require first computing new variables that sum the total values of assets. The following three variables need to be computed:

- Total value of livestock = sum of assets codes 1 through 4.

- Total value of transport assets = sum of asset codes 5 through 9.
- Total value of appliances and electronics = sum of asset codes 10 through 15.

To create output indicators such as the value of televisions no aggregation is needed, however the transfer of data follows the same procedure. To be certain that you can later remember the variable you are creating, click on "Name & Label" at the bottom of the AGGREGATE DATA dialog box. In the VARIABLE NAME AND LABEL dialogue box (Figure 6.3), fill in an identifying name and label for the variable being created.

FIGURE 6.3 SPSS dialogue box for naming and labeling an aggregate variable



6.1.3 Saving the output as new files

In most cases, more than one individual variable can be aggregated at a time and the resulting output indicators saved in the same output file. However, if the aggregation procedure requires that only a subset of cases be selected to complete the aggregation, then separate output files are needed for each aggregation using the SELECT IF procedure. Each output file requires a unique name. The entire aggregation process will require forming nearly a dozen temporary output files. Use file names that will help you remember the contents of each.

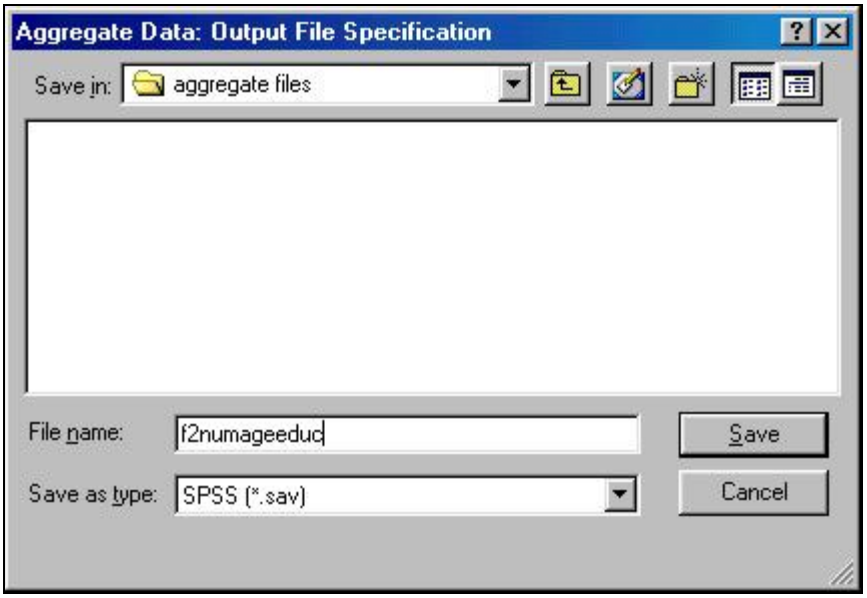
Saving the output for each group of aggregated variables requires making a new file. In the AGGREGATE DATA dialog box, click on the small circle next to "Create new data file" and then click on the "File" button. This displays the OUTPUT FILE SELECTION dialog box (Figure 6.4) is displayed, where the name the file containing the new variables can be entered. Use a name that you will easily recognize at a later time.

6.2 Merging files

In the previous section, guidelines were given for creating many new temporary files containing variables of aggregated data. In SPSS the procedure for merging files is used to combine variables from two different files. Merging different variables for the same cases requires that both files share a common variable with unique values for each case (the household code) and are sorted so that the shared variable is listed in the same sequence in both files (for example, smallest to largest ID code).

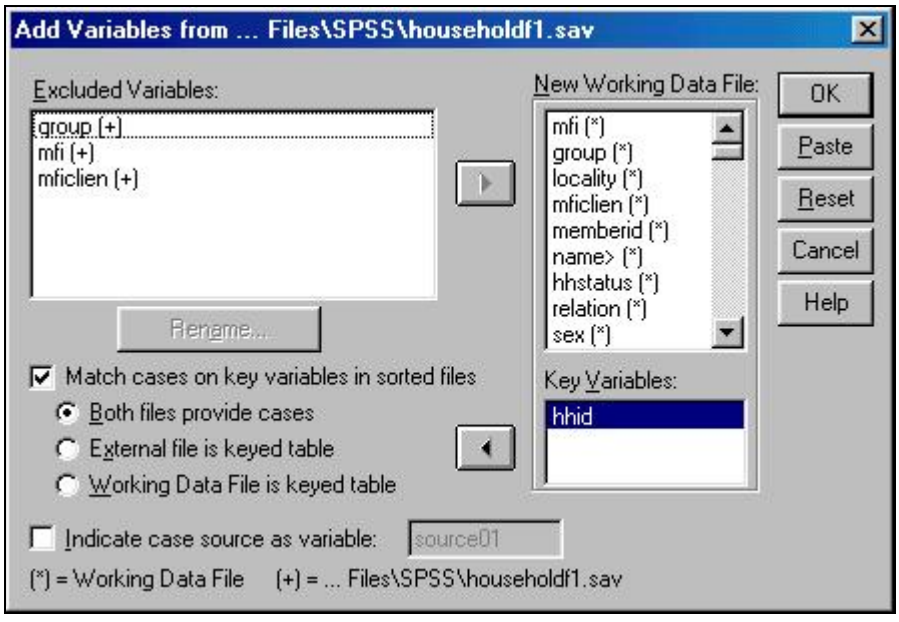
Files can be sorted by selecting **Sort Cases** from the **Data** menu. Select the household ID code variable and move it to the "Sort to" box; then click on "Ascending," followed by "OK." Save the sorted file.

FIGURE 6.4 SPSS dialogue box for outputting file specifications



Once the household file and temporary aggregated data file are sorted by household ID code in ascending order, select **Merge** from the **Data** menu, and then click on **Add Variables**. The ADD VARIABLES FROM: READ FILE dialog box opens . Select the household file to which you want to add the new variables. Click on "Open," and the ADD VARIABLES FROM dialog box (Figure 6.5) opens. SPSS automatically identifies the common variables, which always include the household ID code. Check the box "Match cases on key variables in sorted files," then click on the ID code and move it to the "Key Variables" box by clicking on the arrow button next to it. Click on "OK." Check that the variables have been correctly merged and then save the new file under a different name.

FIGURE 6.5 SPSS dialogue box for merging files



To create a complete file at the household level, all variables listed in Tables 6.1, 6.2, and 6.3 should be aggregated and merged with household-level data.

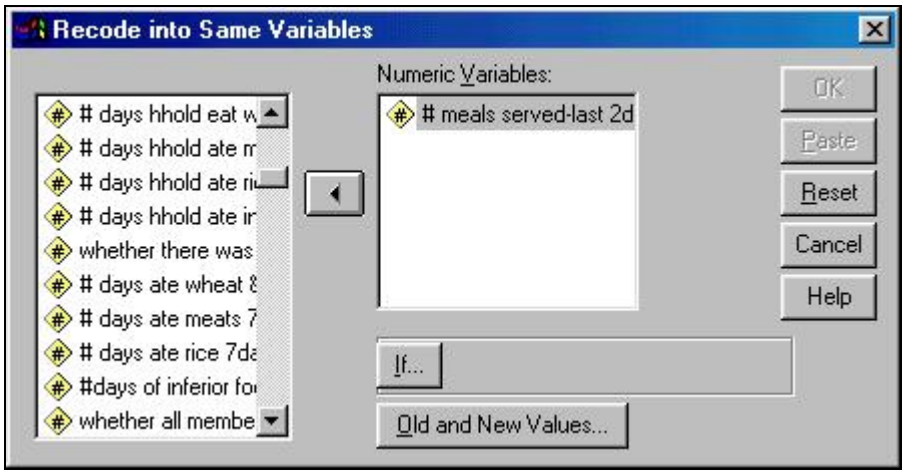
6.3 Transform variables to recode data

You can recode a variable to create a new variable or you can recode a variable and replace the variable that already exists. Recoding is sometimes required to conduct computations. Recoding SYSMIS values into 0 prevents the case from being excluded from a computation or analysis. For instance, to add the responses from C2 to the responses from C3, all sysmis codes can first be changed to 0 and then added to create a new variable measuring the number of meals eaten by all households.

From the **Transform** menu, select **Recode**, then **Into Same Variables**. In the dialogue box displayed specify the variables to be recoded. A selection rule can be specified by clicking on "If."

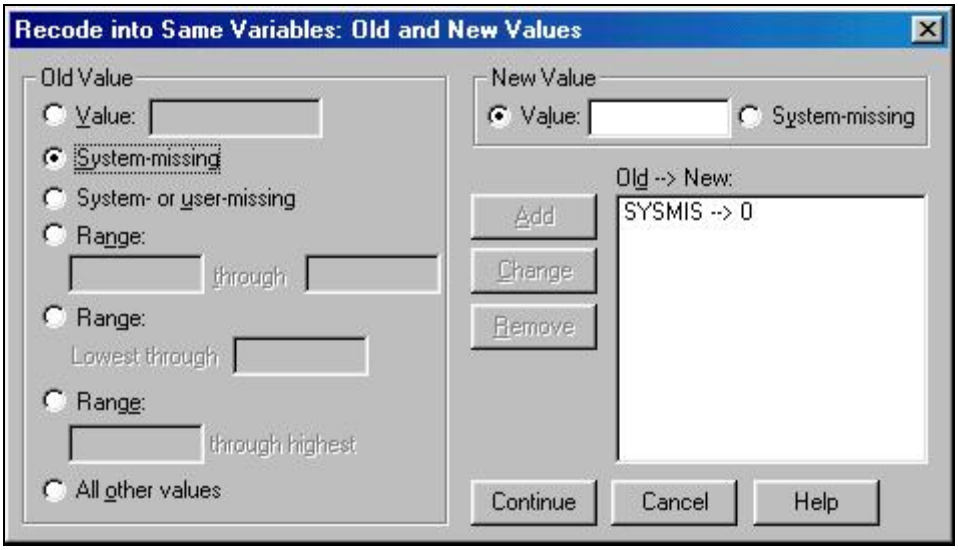
Recode…Into Same Variables reassigns the values of exiting variables or collapses ranges of existing values into new values. Recoding into the same variable can also be used to transform codes missing in the system into another value. Figure 6.6 shows the main dialogue box for recoding into the same variable.

FIGURE 6.6 SPSS dialogue box for recoding into the same variable



To recode into the same variable, click on "Old and New Values" and use the displayed dialog box (Figure 6.7) to indicate which old values are to be changed and what their new values will be. After selecting old and new values, click on "Continue," then "OK," to run the recoding procedure. In Figure 6.7 the value SYSMIS is changed to 0.

FIGURE 6.7 SPSS dialogue box for recoding old values to new values



6.4 Data procedures for computing new variables

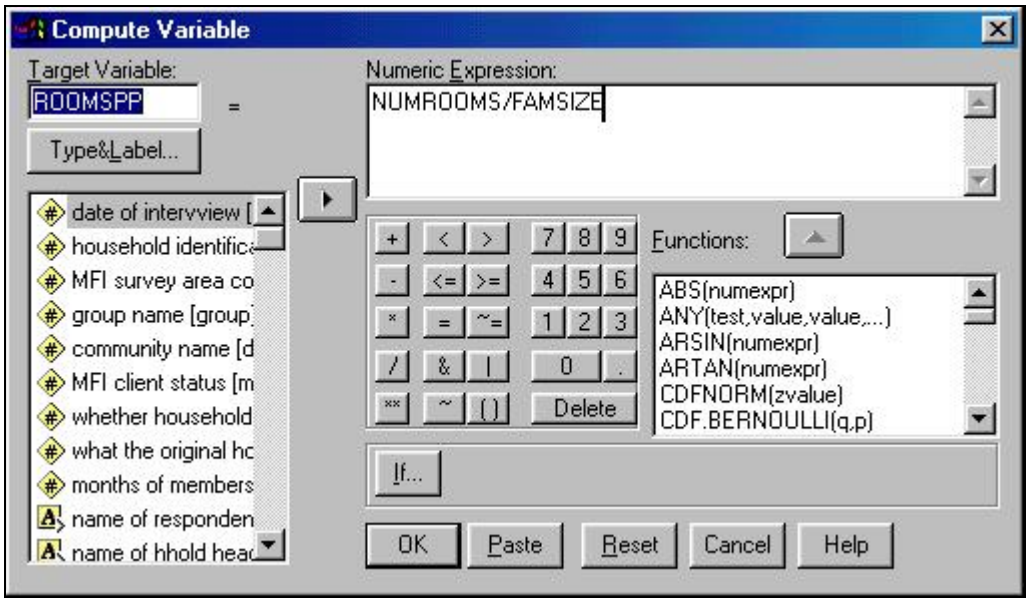
Early in the analysis, the researcher will need to compute new variables from existing variables. These computed variables will be used later in the analysis stage. Table 6.4 shows which computations are needed for creating new variables in the household file.

TABLE 6.4 Computation and output variables at the household level

Variable to be computed	Process for computing variable
Family size (FAMSIZE)	Add the number of adults (NUMADULT) and number of children (NUMCHILD).
Per capita expenditures on food and clothing (PCEXPEND)	Add clothing and footwear expenditures of adults (ADUEXPEN) and children (KIDEXPEN), and divide by family size (FAMSIZE).
Percent of household adults who can write (PRCWRIT)	Divide the number of adults who can write (NUMWRITE) by the number of adults in the household (NUMADULT).
Percent of household adults who completed certain levels of education (PRCEDUC1, PRCEDUC2)	Divide the level of education (for example, EDUC1) by NUMADULT.
Percent of household adults with occupations of different types (PRCOCCU1, PRCOCCU2)	Divide the type of occupation (for example, OCCUP1) by NUMADULT.
Child dependency ratio (CHILDEPN)	Divide NUMCHILD by NUMADULT.
Unemployed dependency ratio (UNEMPDEP)	Divide NUMUNEMP by NUMADULT.
Rooms per person (ROOMSPP)	Divide NUMROOMS by FAMSIZE.
Total number of meals eaten in past two days (NUMMEALS)	Recode missing values to 0, and add MEALS2DY and EVEMEAL2.
Total land owned by household (LANDOWND)	Recode missing values to 0, and add land area cultivated (AREACULT) to land area (AREAUNCU)
Total value of landholdings (VALULAND)	Add the value of cultivated land (VALCULTI) to the value of uncultivated land (VALUNCUL).
Total value of household assets (VALASSETS)	Add together VALANIMA, VALAPPLI, VALTRANS, and VALULAND.
Per person value of total assets (PPASSETS)	Divide VALASSETS by FAMSIZE.

To compute new variables in SPSS, select **Transform** from the main menu, then **Compute**. The COMPUTE VARIABLE dialogue box (Figure 6.8) opens. Click in the "Target Variable" box and type the name of the new variable you are computing. Then click in the **"**Numeric Expression" box and type the variables to be used in computing the new variable. The formula can either be typed in or compiled by clicking on variables from the variable list, followed by clicking on the arrow button, and by clicking on the "Functions" button. Once the variable is created, open the VARIABLE LABEL dialogue box (by clicking on "Type&label") to enter a variable definition and any value labels, if appropriate.

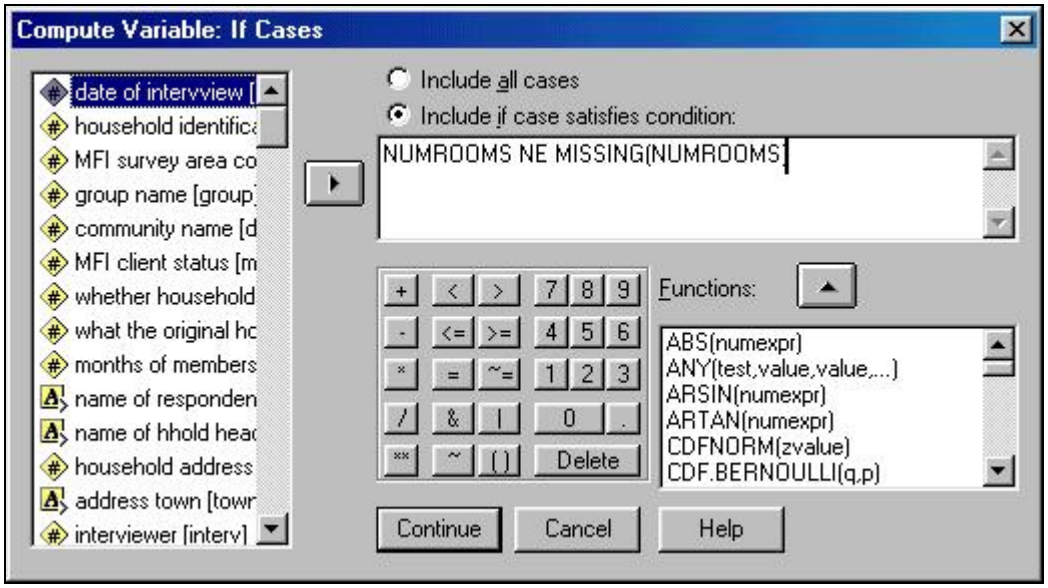
FIGURE 6.8 SPSS dialogue box for computing a new variable



In some cases, computing new variables may require that a condition be applied to filter the values of the existing variable or variables to be used in forming the new variable. SPSS has a special dialogue box (Figure 6.9) for this purpose, which can be accessed by clicking on the "If" button shown in figure 6.8. In the displayed dialog box, a rule for selecting or excluding specific variable cases can be written. When this is completed, click on "Continue," then "OK" to run the computation.

Special precaution should be taken when computing variables from other variables containing missing values (SYSMIS). Any number added to a SYSMIS will result in a SYSMIS. To avoid this problem, exclude them from the calculation through an "if" statement using the expression NE MISSING [variable name] (Figure 6.9).

FIGURE 6.9 SPSS dialogue box for setting the "if" condition to transform data



6.5 Summary

The information contained in this chapter has guided users in aggregating and merging data from the adult (*F2*), children (*F3*) and asset (*F4*) files with data contained in the household file. The chapter also covered guidelines for transforming variables to create new household indicators. The end result of this process was the creation of an expanded household file containing all socioeconomic and poverty indicators required to complete the poverty assessment. This file can now be used to complete the initial stages of analysis to support creation of a poverty index.

CHAPTER SEVEN:
CONDUCTING DESCRIPTIVE DATA ANALYSIS

7.1 Purpose of testing for significant differences between client and nonclient households

Checking for differences between clients and nonclients on a number of socioeconomic characteristics can improve understanding of why differences in poverty levels between the two groups occur. For instance, in a recent study, MFI client households were found, on average, to be younger, better educated, and with fewer unemployed adults, than nonclient households. In two studies of MFIs lending to microenterprise households, client households were found to have a far greater share of members working in business and to have more durable assets. These kinds of details provide background that can later assist in interpreting the assessment's quantitative poverty-related findings. This chapter will provide guidance in how to test for differences between clients and nonclients to identify sampling differences between the two groups.

Differences between groups can be tested using both the *t*-test of differences between means and the chi-square test for cross tabulations. Determining which test to apply depends on the type of data scale used to measure the variable. For nominal and ordinal data, descriptive analysis of the relationship between two variables involves study of cross-tabulation tables to identify patterns of responses that differ by client status. To test whether differences in responses between sample groups are significant—that is, the variables are not independent of each other—the chi-square test is used.

The *t*-test of differences between two means is used to determine whether the difference in means between two groups of independent samples for an interval variable is significant. Significant differences found in the samples can be interpreted as representative of the population. In this case, the population refers to the entire group of MFI new clients and nonclients located in the same area.

7.2 Cross tabulation and chi-square test

7.2.1 How cross tabulation is applied

When one or more variables are measured on a nominal or ordinal scale, cross tabulation is a means of identifying a relationship between two variables. Cross tabulation categorizes into cells the number and percent of cases in which different combinations of responses occur. For example, if the researcher wants to check for differences in the principal occupations for surveyed client and nonclient households, a cross tabulation would show the absolute numbers of the two household types working in each category of occupation, as well as the percent of total households falling in each category. In this way, patterns that differ between clients and nonclients can more easily be detected. Table 7.1 shows the cross tabulation segmenting responses by client status and occupation types for recent case study. As the table shows, on a percent basis, nonclients are more likely to be

self-employed in agriculture, work in casual labor, or be unemployed and looking for a job. MFI clients are more likely to be self-employed in a nonfarm enterprise. These results are noteworthy because they indicate that MFI selection criteria such as the client being engaged in microenterprise activity translates into an overrepresentation of households with adults engaged in business than would be expected in the general population. If households engaged in business tend to have higher or lower poverty levels than nonbusiness households, this would likely influence the overall ranking of client households within the poverty index.

TABLE 7.1 Crosstabulation of client status by main occupation of adults in household

Main occupation		client	nonclient	Total
self employed in Agriculture	Count	125	311	436
	percent within column	17.5%	32.3%	26.0%
Self employed in nonfarm enterprise	Count	218	151	369
	percent within column	30.6%	15.7%	22.0%
pupil/student	Count	154	148	302
	percent within column	21.6%	15.4%	18.0%
casual	Count	20	59	79
	percent within column	2.8%	6.1%	4.7%
salaried worker	Count	98	143	241
	percent within column	13.7%	14.8%	14.4%
domestic work	Count	62	75	137
	percent within column	8.7%	7.8%	8.2%
unemployed, looking for a job	Count	22	62	84
	percent within column	3.1%	6.4%	5.0%
unwilling to work/retired	Count	6	8	14
	percent within column	.8%	.8%	.8%
not able to work/handicapped	Count	8	6	14
	percent within column	1.1%	.6%	.8%
	Count	713	963	1676
	percent within column	100.0%	100.0%	100.0%

Determining whether differences in the distribution of responses across categories are significant in a statistical sense is done through the application of the chi-square test. The chi-square test answers the question of whether the observed differences in responses between categories reflect sampling error or indicate a relationship. In the example of the occupation of household adults, a chi-square value that is significant at 0.05 (or less) suggests that a difference between client and nonclient households exists in terms of the distribution of occupation. The nature of this relationship, however, can only be discovered through inspection of the cross-tabulation table.

Table 7.2 shows the chi-square results for the cross tabulation shown in Table 7.1. The chi-square level of significance is less than 0.001, indicating that a very strong difference exists in the pattern of occupation responses between clients and nonclients.

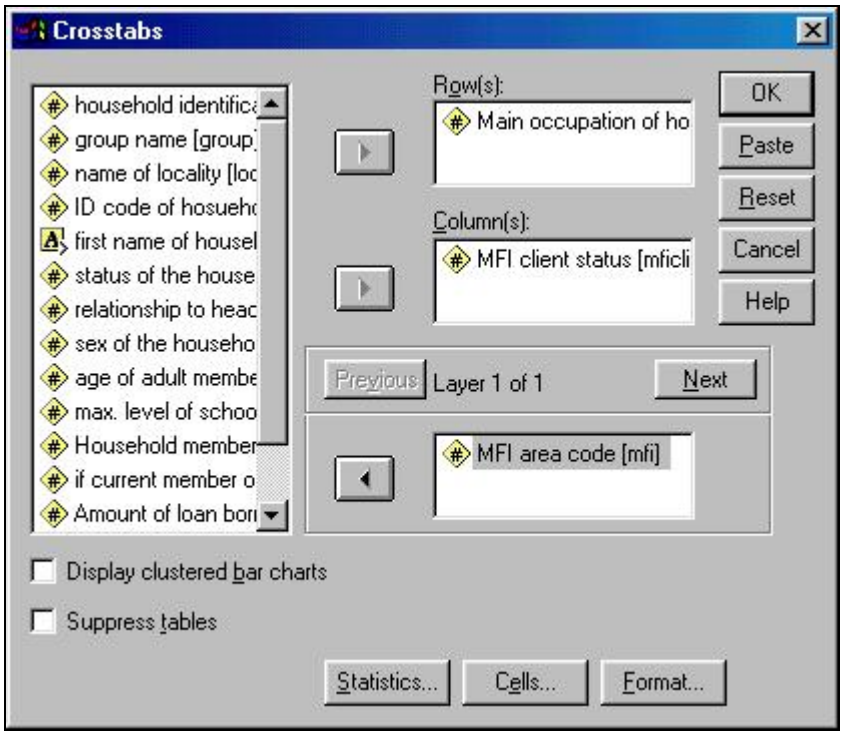
TABLE 7.2 SPSS output table for Chi Square test of cross tabulation

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	105.191	8	.000
Likelihood Ratio	107.098	8	.000
Linear-by-Linear Association	.022	1	.881
N of Valid Cases	1676		

7.2.2 The cross-tabulation procedure in SPSS

To run the cross-tabulation procedure in SPSS, click on **Descriptive Statistics** in the **Analyze** menu, then the option **Crosstabs**. This will open the Crosstabs dialogue box (Figure 7.1). Move the variable for designating client status from the list at the left to the "Column(s)" box. Move the variables to be compared by client status into the "Row(s)" box. More than one variable can be selected at a time in the "Row(s)" box, but each combination will result in a separate cross-tabulation table. To run a cross tabulation for each survey cluster, mover the MFI cluster code to the "Layer 1 of 1" box.

FIGURE 7.1 SPSS dialogue box for running cross tabulations



Now click on "Statistics" at the bottom of the page to open the Statistics dialog box (Figure 7.2). Check the box for **Chi Square** and click OK to return to the Statistics: Crosstabs dialogue box. Click on the "Cells" to open the Crosstabs: Cell Display dialog box (Figure 7.3) and check the boxes for "Counts…Observed" and "Percentages…Column." Click on "OK" to return to the Crosstabs

dialogue box, then on “OK” there to run the cross tabulation. The results automatically appear in the SPSS OUTPUT VIEW window.

FIGURE 7.2 SPSS dialogue box for cross-tabulation statistics

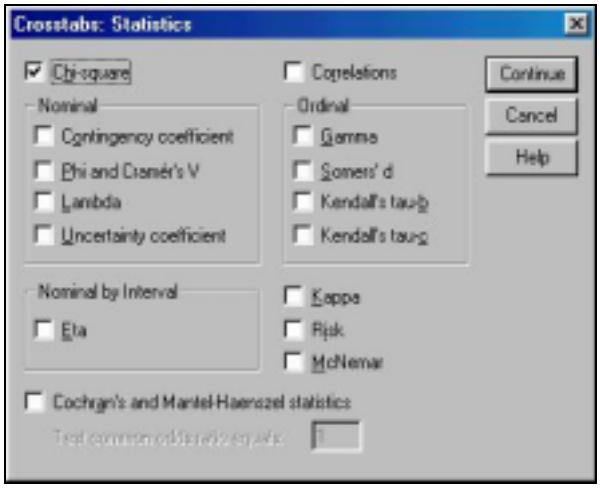
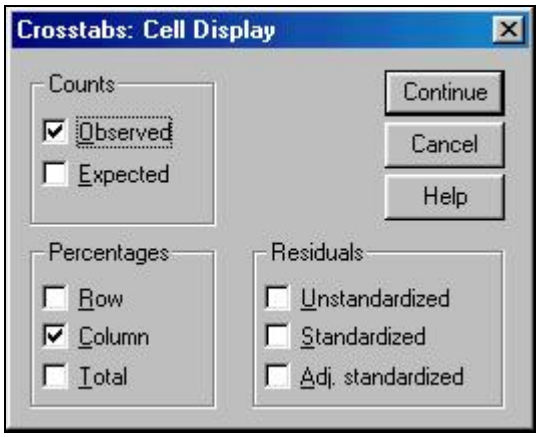


FIGURE 7.3 SPSS dialogue box for setting types of cell display in the cross-tabulation table



7.2.3 Interpreting a cross-tabulation table

When running cross tabulations, it is possible to enter multiple variables into the row at once, but be certain to identify only the variable for client status in the column text box, as shown in Figure 7.1. Check only the option for showing percentages for columns, as the row percentage breakdown will not be useful. Once the test is run, check the results in the output file to screen for any specification errors.

Interpreting a cross-tabulation table accurately takes practice. In most cases, the absolute number given in each cell of the table provides little insight. Instead, noting the percent of the total number of cases falling into each cell can make differences in the distribution easier to see. Percentages can be given as either a share of the total number of cases in the column, or the total number of cases in

the row. In this study, percentage breakdowns of column totals will most often be used, where the column variable indicates whether the respondent is an MFI client household or nonclient household.

Cross tabulation can be done at different levels of data. Further clarification into the pattern of differences between clients and nonclients may be gained by dividing data into smaller categories, such as individual survey regions. In this way, the source of the differences showing up in the aggregated data may be pinpointed to a more specific relationship. At the same time, cross-tabulation that is more detailed may uncover a spurious relationship that appears only when subcategories are aggregated. The question to ask when deciding how to delve deeper is "Under what circumstances does this relationship exist?"

In the previous example shown in Table 7.2, a test of significance at the cluster level for the same data set uncovered the results shown in Table 7.3. Based on the chi-square levels of significance, we note that occupational differences between clients and nonclients exist in four of the five regions. The region where no differences are found is highly urban with fewer opportunities for agricultural enterprises.

TABLE 7.3 Tests of significance between client status and occupation at cluster level

Cluster code		Value	df	Asymp. Sig. (2-sided)
Cluster 1	Pearson Chi-Square	30.970	6	.000
	Likelihood Ratio	31.864	6	.000
	Linear-by-Linear Association	23.413	1	.000
	N of Valid Cases	379		
Cluster 2	Pearson Chi-Square	13.707	6	.033
	Likelihood Ratio	14.041	6	.029
	Linear-by-Linear Association	.616	1	.433
	N of Valid Cases	336		
Cluster 3	Pearson Chi-Square	1.043	6	.984
	Likelihood Ratio	1.045	6	.984
	Linear-by-Linear Association	.156	1	.693
	N of Valid Cases	288		
Cluster 4	Pearson Chi-Square	18.361	6	.005
	Likelihood Ratio	19.178	6	.004
	Linear-by-Linear Association	8.979	1	.003
	N of Valid Cases	342		
Cluster 5	Pearson Chi-Square	15.539	6	.016
	Likelihood Ratio	15.833	6	.015
	Linear-by-Linear Association	4.266	1	.039
	N of Valid Cases	331		

7.2.4. Conducting specific analysis using cross tabulations.

- The following list of indicators can be used to test for significant differences between clients and nonclients using cross tabulation. The analysis for each indicator can also be repeated at the cluster level.
- From the adult file (*F2*):

- Main occupation of household adults
- Education levels of household adults
- Marital status of household head

If significance differences are found between the occupation and levels of education, interpret the likely source of these differences.

7.3 The *t*-test on the difference between means

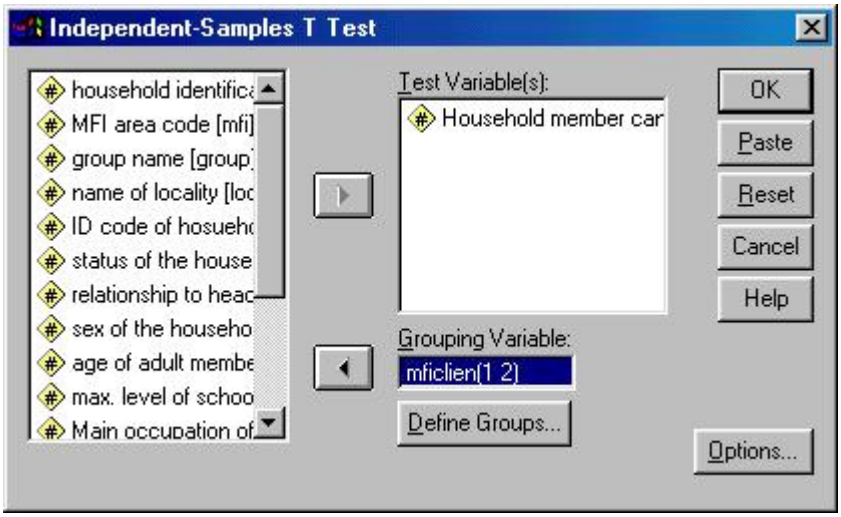
7.3.1 How the t-test is applied

For most socioeconomic indicators in this assessment tool, the number of possible values for a variable will be too large to make use of a cross-tabulation table. This is particularly the case for interval- and ratio-scaled variables. One way to test for significant differences between MFI clients and nonclients on interval and ratio data is to compare the means of a variable for the two different groups. Here, the mean differences between the two groups and the deviation from the mean within each group are used to derive a *t*-value. This value can then be compared with what is called the "critical *t*-value" and, if higher, the groups can be considered different. On the other hand, if the calculated *t* is lower than the critical *t*-value, one can conclude that no difference exists between the two groups regarding the variable in question. If the actual t-value is above the critical *t*-value, the level of significance will be .05 or less.

7.3.2 SPSS procedure for running a t test of means

To run a *t*-test, click on **Compare Means** in the **Analyze** menu, then on **Independent-Samples *T* Test**. This opens a dialogue box (Figure 7.4) where interval or ratio variables can be selected as test variables. The grouping variable, client status, identifies how to differentiate two groups of cases. Select only the MFI client status as the single grouping variable, and then click on "Define Groups" to specify two codes for the groups to be compared. Be certain that the codes used match those entered in the data file.

FIGURE 7.4 SPSS dialogue box for running an independent *t*-test on samples



Tables 7.4 and 7.5 shows SPSS output tables for comparing the mean of two independent samples. The first table shows the number of cases from each subcategory used for the calculation. The middle columns in this table show the calculated means for the two groups of MFI clients and nonclients and the standard deviation associated with each mean. The table shows the share of households adults who can write by client status. Because the underlying code used 0 for no and 1 for yes, the resulting means can be easily translated into percent. Ninety-five percent of adults in client households can write compared with ninety-two percent of adults in nonclient households.

TABLE 7.4 SPSS output table for independent *t*-test on samples

	N	Mean	Standard deviation	Standard error mean
Adults who can write				
Client of MFI	713	.95	.22	8.42E-03
Nonclient of MFI	963	.92	.27	8.80E-03

Determining whether the means are significantly different requires studying the second output shown in table 7.5. Here the table indicates first whether the variances between the two groups can be considered equal (Levene's test). If the level of significance is less than 0.05, the calculated *t*-value is that shown in the row for equal variance. The calculated *t*-value for this example is 2.2 and the significance of this value is 0.03, indicating that the calculated *t*-value is significantly greater than the critical *t*-value. On the basis of this result, one can conclude that MFI households have a significantly greater percent of adults who can write than nonclient households.

TABLE 7.5 Independent samples t-test

	Levene's test for equality of variances		***t*-test for equality of means**				
	F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean difference**	**Std. error difference**
Adults who can write							
Equal variances assumed	19.96	0.000	2.211	1674	.027	2.77E-02	1.25E-02
Equal variances not assumed			2.275		.023	2.77E-02	1.25E-02

7.3.3 Conducting specific analysis using t-test of means.

Results of interval- and ratio-scaled data tested using the *t*-test of means can be summarized in an SPSS output file. In addition, a summary narrative sheet can be prepared describing significant differences between clients and nonclients found by analyzing the following variables.

Expanded household file (*F1*):

- Family size
- Number of children
- Percent of female-headed households
- Average size of landholdings
- Average value of landholdings

Adult file (*F2*):

- Percent of adults who can write

Children's file (*F3*):

- Average age of children

7.4 Summary

In this chapter, analysis has focused on how to identify differences between clients and nonclients based on a number of socioeconomic indicators. When differences between the two groups are found to be significant, this information may suggest that selection criteria of the MFI has resulted in the sampled groups being different in ways that are not directly related to their poverty status but which could influence their status. These differences should be noted when interpreting results from measurement of the relative poverty of households. Measurement of household relative poverty will be explained in the next chapter.

CHAPTER EIGHT: DEVELOPING A POVERTY INDEX

A basic premise of this poverty assessment tool is that, within the range of poverty indicators collected through survey techniques, a subset of indicators exists that measures different aspects of relative poverty at the household level. Which combinations of indicators prove the most instrumental in measuring relative poverty in a given survey area will differ, and often in ways that are somewhat predictable. In countries where poverty is extreme, indicators signaling chronic hunger tend to differentiate the relative poverty of households. In densely populated countries, ownership of land and dwellings may better signal differences in relative poverty. Cultural differences will also influence the strength of some types of indicators.

Developing an objective measure of poverty then requires first identifying the strongest individual indicators that distinguish relative levels of poverty and then pooling their explanatory power into a single index. This chapter guides users in conducting analysis of data to:

1. determine which indicators are the strongest measures of relative poverty for the surveyed households,
2. create a ranking list of these variables on the basis of correlation levels with the poverty benchmark indicator—per capita expenditures on clothing and footwear, and
3. apply these ranked indicators systematically to calculate a household poverty index.

8.1 Statistical procedures for filtering the strongest indicators for measuring poverty

8.1.1 Linear correlation coefficient

The linear correlation coefficient procedure is the primary means of filtering poverty indicators to determine which variables appear the strongest in capturing differences in the relative poverty of households. Testing the level and direction of correlation among a wide array of ordinal and interval variables with the benchmark poverty indicator—per capita expenditures on clothing and footwear (PCEXPEND)—is the primary means of determining the strength of indicators.

The linear correlation coefficient is a statistical procedure used to measure the degree to which two variables are associated. The correlation coefficient can determine the level and direction of a relationship between two variables. Linear correlation does not require that the units used in each variable be the same. The values of the correlation coefficient range from –1.00 to +1.00, and their sign and magnitude indicate how the two variables relate to each other. A coefficient value at or near –1 indicates that the variables are inversely related, or that a higher value for one is associated with a lower value for the other. Higher education may be inversely related to consumption of inferior food,

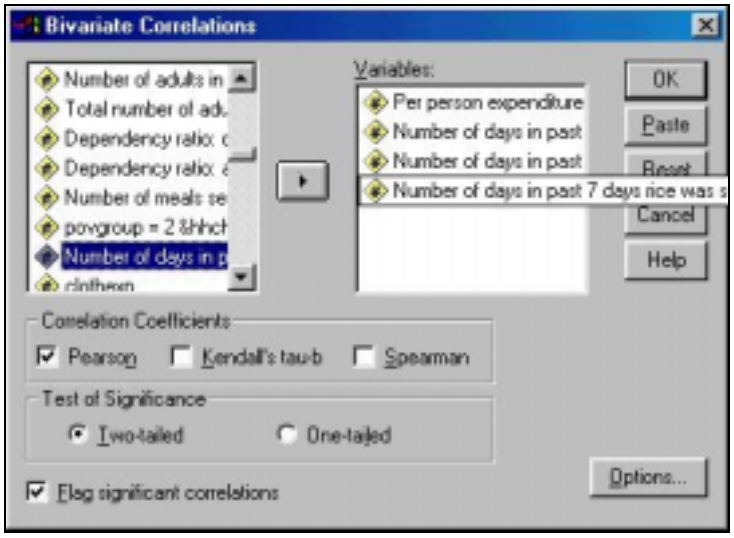
since higher education often brings higher income—which in turn pays for better-quality food. In contrast, a value at or near 1 suggests a strong positive relationship between the two variables. For example, the number of household members may be very closely related to the number of rooms in the household. Coefficient values at or near 0 suggest that no strong relation exists between the variables.

The interpretation of results is based on probability theory that determines the level of significance with which differences among sample groups can be applied to the entire survey population. In the assessment tool, levels of significance are set at 0.10 or less, meaning that a minimum 90-percent confidence interval is used to either accept or reject the hypothesis that the association between two variables is random. If the level of significance is found to be less than 0.10, one can conclude that some association between the variables exists. If the significance level is found to be less than 0.05, then the association is strong, and if the significance level is found to be less than .01, then the association is considered very strong.

8.1.2 Using SPSS to measure linear correlation

Correlation tables are created in SPSS by selecting **Correlations** under the **Analyze** menu, then **Bivariate** as the type of correlation. This will open the BIVARIATE CORRELATIONS dialogue box (Figure 8.1). Highlight variables in ordinal, interval or ratio form in the variable list at the left and move these to the text box at the right by clicking on the arrow button. At the bottom of the dialogue box, check "Pearson" as the type of correlation coefficient, and select "Two-tailed" as the type of significance. Choose as a first variable in the center dialogue box the variable PCEXPEND, or per capita expenditure on clothing and footwear. Add additional variables from the list of indicators shown in Figure 8.2 in groups of not more than six to eight at a time.

FIGURE 8.1 SPSS dialogue box for bivariate correlations



8.1.3 Interpreting the SPSS correlation table

The correlation table created in SPSS is in matrix form and will appear in the OUTPUT VIEW window. If the variable PCEXPEND tops the list in the BIVARIATE CORRELATIONS dialogue box, the first column in the output table shows the levels of correlation between PCEXPEND and all other variables run in the procedure.

FIGURE 8.2 Ordinal- and interval-scaled indicator variables

Human Resources (Section B)
Average age of household adults
Percent of adults who can write
Percent of adults who completed specific levels of education
Percent of adults with specific occupation
Number of children
Family size
Dependency ratio of children to adults
Dependency ratio of unemployed to employed
Per person expenditure on clothing and footwear
Food Security & Vulnerability (Section C)
Number of meals in past 2 days
Number of days when luxury food 1 served
Number of days when luxury food 2 served
Number of days when luxury food 3 served
Number of days when inferior food served
Number of days not enough food in past month
Number of months not enough food in past year
Weeks of stock for food staple
Frequency of purchase staple 1
Frequency of purchase staple 2
Frequency of purchase staple 3
Share of additional income spent on food
Dwelling (Section D)
Number of rooms per person
Structural condition of house
Quality of latrines
Quality of drinking water
Quality of dwelling walls
Quality of roofing
Quality of floors
Extent of electrical use
Quality of cooking fuel
Ownership status of house
Assets (Section E)
Value of appliances and electronics
Value of transport
Value of landholdings (irrigated and unirrigated)
Quantity of land owned
Value of animals
Value of assets per person (or per adult)

Table 8.1 is an example of a correlation output table. The results shown indicate that, of the three variables correlated with per person clothing expenditures (the shaded boxes), only the first two are found to be significantly associated. These are number of days meat and rice were eaten, and each was found significant at less than $p = 0.01$, indicating a 99-percent certainty that the correlation is not random. Note that, although the correlation coefficient for days rice eaten is 0.179, the correlation is still considered highly significant. The variable for days inferior foods are eaten is negatively correlated with expenditures, as would be expected, but the level of significance (0.376) indicates that no association exists between per capita expenditures on clothing and footwear and consumption of inferior food.

Using the output shown in Table 8.1, the two variables for number of days meat and rice were eaten can be added to a filtered list of indicators measuring aspects of poverty. To complete the filtering process, all other variables listed in Figure 8.2 are correlated with PCEXPEND and those registering a significant level of correlation are added to the filtered list of poverty indicators.

TABLE 8.1 SPSS output table of correlation between indicator variables

Indicator variable		Per person expenditure on clothes and footwear	Number of days rice was served	Number of days meats were served	Days inferior food was served
Per person expenditure on clothes and footwear	Pearson correlation	1.000	0.179	0.439	–0.040
	Sig. (2–tailed)		0.000*	0.000*	0.376
	N	500	499	499	495
Number of days in past 7 days rice was served	Pearson correlation	0.179	1.000	0.328	–0.129
	Sig. (2-tailed)	0.000*	.	0.000*	0.004*
	N	499	499	499	495
Number of days in past 7 days meats were served	Pearson correlation	0.439	0.328	1.000	–0.144
	Sig. (2-tailed)	0.000*	0.000*	0	0.001*
	N	499	499	499	495
Number of days in past 7 days inferior food was served	Pearson correlation	–0.040	–0.129	–0.144	1.000
	Sig. (2-tailed)	0.376	0.004*	0.001*	.
	N	495	495	495	495

* Correlation is significant at the 0.1 level (2-tailed).

In large data sets, such as in this methodology, even small correlation coefficients may also signal an association between two variables. To verify this, check that the association is found significant (level of significance is less than 0.10).

8.1.4 Selecting variables to test for correlation

The correlation procedure should always be set up with the benchmark indicator, per capita expenditures on clothing and footwear, being the first listed in the bivariate correlation dialogue box. To keep the output tables a manageable size, run separate correlation tables for each group of indicators listed in Figure 8.2. By always including PCEXPEND as the first variable listed, the first column of the output will always show the correlation coefficients between the benchmark poverty indicator and all others included. The list of variables in Figure 8.2 includes all ordinal- and interval-scaled variables that can be correlated against the variable PCEXPEND.

Output from the analysis can be summarized in a table listing all indicators tested and ordered according to the strength of association measured, and noting the number of cases found with missing values. An example of this format is shown in Table 8.2. Indicators registering the highest levels of significance ($p < 0.10$) would top the list, while indicators registering insignificant levels of association, with $p > 0.10$, would be excluded from the list. It is also important to note the sign of the

correlation coefficient to remember if the relationship was found to be a negative or positive one. This table will be used again when estimating the poverty index.

TABLE 8.2 Template for recording ranked indicators by level of association with benchmark poverty indicator, per person expenditures on clothing and footwear

Indicator	Level of significance	Sign of the correlation coefficient	Number of cases with missing values
1.			
2.			
3.			
…			

8.2 Principal component analysis to estimate a poverty index

This assessment tool develops a relative poverty index through the application of principal component analysis (PCA). The PCA method is applied to determine how information from various indicators can be most effectively combined to measure a household's relative poverty status. The end result of PCA is the creation of a single index of relative poverty that assigns to each sample household a specific value, called a score, representing that household's poverty status in relation to all other households in the sample. The analyst creates the index from the combination of individual indicators that significantly correlate with one another on the basis of a shared underlying poverty component.

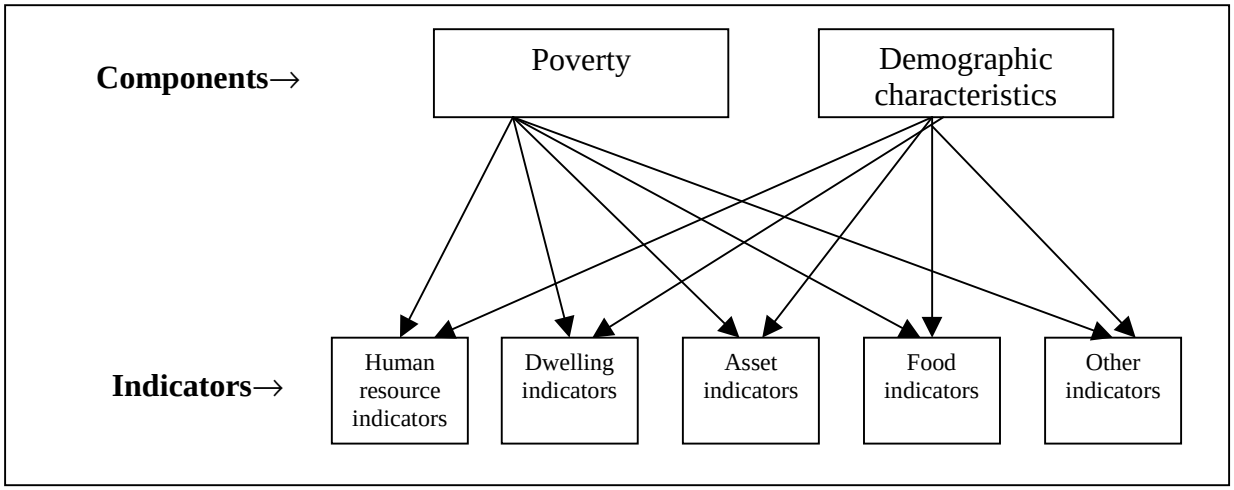
PCA is used to identify, or extract, within a group of indicators, underlying components that can at least partially explain why the indicator values differ between households in the way they do. Each component is assumed to capture a unique attribute shared by households. One of the reasons why households answer differently to indicator questions is because of their relative poverty status.

If the indicators are related in more than one way, then more than one underlying component will be created. However, only one component will measure a household's relative poverty. Indicators may also relate to one another because of the rural or urban setting of households or specific regional conditions. Other possible underlying components may capture aspects related to similarities between households in education, occupation, or cultural practices. In general each component extracted will capture a unique attribute shared by households. Further, the number of components that can be "extracted" increases with the number of indicators included in the analysis. Figure 8.3 shows how components relate to the indicator variables used to describe them.

The principle objective of using PCA in the poverty assessment exercise is to extract the "poverty component" that can be used to compute a household specific index of relative poverty. Hence, the application of PCA will use first and foremost indicators that already show strong correlation with the poverty benchmark indicator, per person expenditure on clothing and footwear. Filtering the indicators in this way will support a stronger measure of a poverty component – one that associates most consistently and strongly with what the analyst expects to closely measure relative poverty.

This component can then be treated as a "poverty index." The following sections guide users in how PCA method is applied to most effectively measure the poverty index.

FIGURE 8.3 Indicators and underlying components



8.3 Statistical tools used in creating a poverty index

The steps for creating a poverty index using the PCA method are as follows:

- Select a screened group of variables correlated with the poverty benchmark indicator
- Run a test model and interpret the results
- Revise the model on the basis of the results of prior runs until the results meet the performance requirements
- From a final model, save poverty component scores as poverty index variables.

8.3.1 Step one: Select a screened group of indicators correlated with the poverty benchmark indicator

Before the PCA method is applied to the data, poverty indicators must go through a series of filters to ensure that the resulting index does not represent a distorted measure of poverty. In section 8.1, a list of all indicators correlated with the poverty benchmark indicator, per person expenditure on clothing and footwear, was created. The reduced list of indicators in Table 8.2 constitutes the first screening of indicators for PCA. These indicator variables are all in ordinal, interval and ratio scale, which is required for the PCA method. Check the list for any variables with more than 25 missing values and use these as sparingly as possible.

The following additional filters are used to further narrow selection of variables for the PCA model:

- Limit the number of indicators used in PCA. Having fewer variables reduces the complexity of the resulting calculated components. Closely related variables that effectively measure the same phenomenon can be screened, with only the strongest added to the PCA model. For example, if all three luxury foods correlate strongly with per capita expenditure on clothing and food, choose only one or two of these. We recommend that no more than 20 variables be used to create the poverty index.
- Balance the range of indicators to reflect different dimensions of poverty. Several indicators measuring similar aspects of poverty can be included in a PCA model; however, a heavy concentration of similar indicator types can inappropriately skew the resulting poverty index to overemphasize one aspect of poverty. To avoid this, select several indicators from each section of the questionnaire.

8.3.2 Step two: Running a test model and interpreting results

Components can be extracted from a series of indicators using several different techniques; however, only one—the principal components technique—is appropriate for the poverty assessment methodology. In PCA, each calculated underlying component represents a linear combination of the indicator variables used in the model.

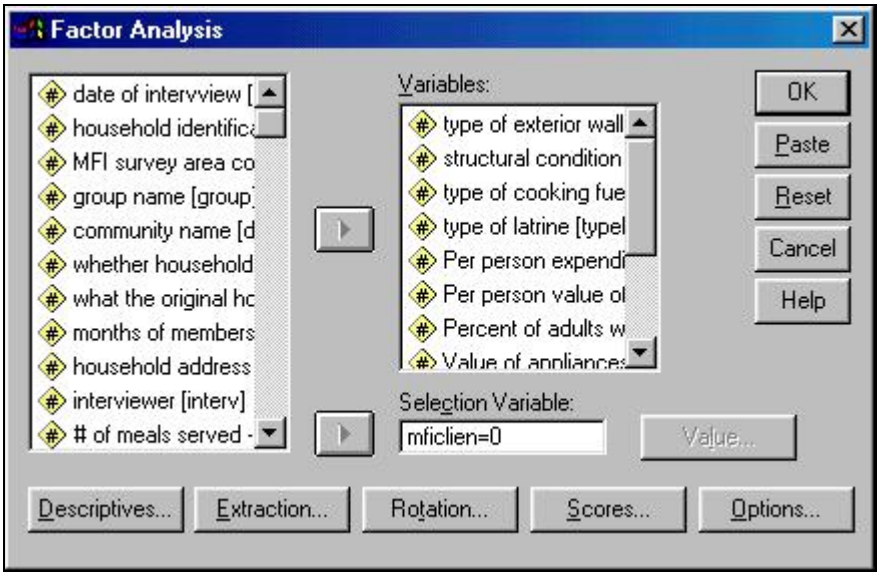
The first component is the combination that accounts for the largest amount of variance in the sample. The second component accounts for the next-largest amount of variance and is uncorrelated with the first. Successive components explain progressively smaller portions of the total sample variance. All components are uncorrelated with each other. Because of this trait, only one can be considered to measure relative poverty.

The SPSS procedure of PCA

You are now ready to run an initial PCA model. From the **Analyze** menu, select the **Data Reduction** option, then **Factor Analysis.** This opens the FACTOR ANALYSIS dialogue box (Figure 8.4). Begin by selecting, from the list of variables correlated with the benchmark poverty indicator (per person expenditures on clothing and footwear), 6 to 10 indicators that register the strongest levels of association (Table 8.2). Scroll down the list of variables to the left and select the highest ranked indicators. Move them to the box on the right by clicking on the upper arrow button.

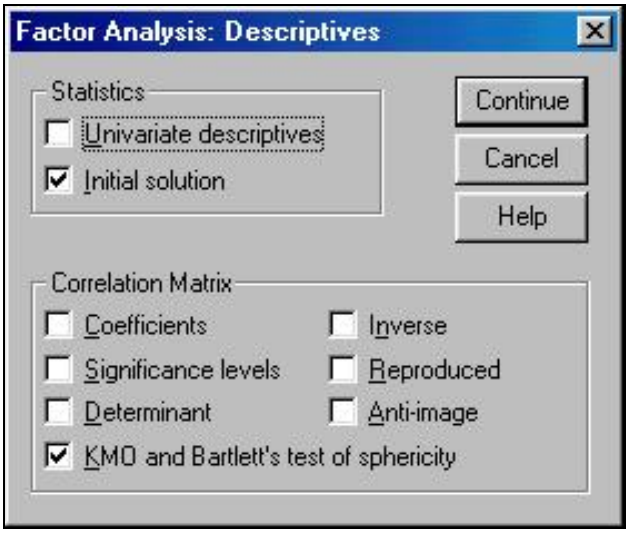
Once you have selected the variables, select from the list the indicator that distinguishes MFI clients from nonclients. Click on the lower arrow button to move this indicator to the SELECTION VARIABLE box. Click on "Value" to choose the value representing nonclient households (designated as "0" on the questionnaire). This will restrict your initial model to include *only the 300 nonclient households*. The nonclient sample represents the general population and is therefore a more appropriate group to use for building the initial model.

FIGURE 8.4 SPSS dialogue boxes for analyzing factors



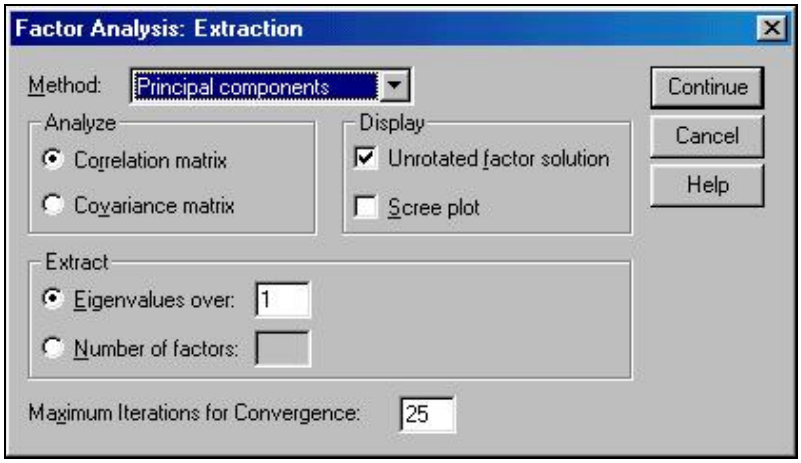
Click on **Descriptives** to open the dialog box shown in Figure 8.5. In this dialogue box, check the box "Initial Solution" in the top part and "KMO and Bartlett's test of sphericity" in the bottom part. Click on "Continue" to return to the main FACTOR ANALYSIS dialogue box.

FIGURE 8.5 SPSS dialogue box for setting statistical tests



In this dialogue box ("Factors Analysis"), click on "Extraction" to open the dialog box shown in Figure 8.6. Set the extraction method to PCA by selecting "Principal components" from the "Method" drop-down list. Under "Analyze," check the box "Correlation Matrix." Also, under "Display," check the box "Unrotated factor solution." Note that, in the lower part of the form, you can alter the minimum value of the Eigen value or limit the number of factors to be extracted. Select the minimum value of the Eigen to 1 (the default value). You will use the latter when saving the final results of the model. Click on "Continue" to return to the main FACTOR ANALYSIS dialogue box.

FIGURE 8.6 SPSS dialogue box for extracting data to PCA



At the bottom center , click on "Rotation." In the dialogue box that opens, check the box for "No Rotation." Click on "Continue" to return to the FACTOR ANALYSIS dialogue box.

Finally, in the lower-right corner dialogue box, click on "Options." In the dialogue box that opens, under "Missing Values," select "Replace with Mean Values." In the bottom section of the form, note the option to sort component coefficients by size, and to choose a minimum value to be shown on the screen. These options can be used later when refining the model. Click on "Continue" to return to the main FACTOR ANALYSIS dialogue box.

Identifying and interpreting the poverty index component in the SPSS output

The initial output for the PCA model will include four tables: the component matrix, the explained common variance table, the communalities table, and the KMO-Bartlett test. Each output can be used to interpret results and refine the model. However, the most critical to determining the indicator composition of the poverty index model will be the output from the component matrix (the fourth table of the analysis—see Table 8.3 for an example). Results shown in other tables may indicate that changes are needed, but the component matrix results will indicate what these changes should be.

Determining how well the PCA model works in creating the poverty index involves assessing the coefficients for each component, called the "component loadings." In fact, analysis of component loadings is the most important determinant for developing the poverty index. Component-loading coefficients represent the amount of correlation between the component variable and the indicator variable. To check whether the PCA model is misspecified, do the following:

1. *Check the size of the absolute value of the coefficients for each indicator.* This indicates the degree of correlation between the component and the indicator. Large absolute values indicate a high level of correlation, while low numbers indicate a lower level of correlation. For a sample size of 500, to be considered significant at the 0.01 level, a factor coefficient should have a minimum value of 0.180 (following the Burt-Banks formula) but are best screened for those above 0.300.

TABLE 8.3 Component matrix

Variable	Component loadings	
	1	2
Number of days in past 7 days wheat was eaten	0.729	–0.109
Number of days in past 7 days meat was eaten	0.772	–3.002E-02
Extra monthly income spent on food	–0.539	0.649
Members had enough to eat in past month	0.512	0.339
Household source of cooking fuel	0.462	0.497
Household electricity use	0.624	–0.342
Quality of latrine	0.560	–0.118
Percent of adults who can write	0.471	0.655
Percent of adults who completed secondary school	0.612	–8.014E-02
Variable	**Component loadings**	
	1	**2**
Per person expenditure on clothes and shoes	0.713	–0.128
Per person value of assets	0.464	-0.244
Value of radios owned by household	0.612	-0.131
Aggregate value of all appliances and electronics	0.654	–0.122

Extraction method: principal components analysis.

2. *Check that the sign of each component coefficient is what would be expected for each indicator in the model.* In other words, the coefficient for all indicators expected to decline as wealth increases should carry a negative sign. Positive coefficients indicate a direct relationship between the indicator and the relative wealth of the household. As the values of an indicator increase, so does the value of the component, which in this case is the relative wealth of the household. Negative coefficients indicate an inverse relationship between the indicator and the relative wealth of the household. Table 8.3 shows the component loading coefficients for a PCA model used to calculate a poverty index. As the table shows, two components were calculated from the indicators.

In the above example, the size of the absolute value of all component loadings on the first component indicate that all can be considered significant explanatory indicators. The signs on the coefficients also align with the expected characteristics of relative poverty. The second component in this model captures another common aspect of households and may suggest a relationship hinging more on rural households. It does not appear to consistently capture variance related to relative poverty, since for some variables the loadings carry an unexpected sign, their magnitude is insignificant, and the results do not appear consistent from one variable to the next.

The analyst can improve the model's explanatory power by screening out variables that have low component loadings on the poverty component, since these do not improve the explanatory power of the index, and by adding new variables from Table 8.2 to see if the addition improves or weakens the model results.

Table 8.4 shows a second component matrix but this time with a few indicators that appear to contribute little to the model. The indicators 'number of days inferior food was served' and 'number

of bulls and cows' have much lower coefficients than others in the model, although the signs of the coefficients are what is expected. (Both of these indicators were found insignificantly correlated with the benchmark poverty indicator and are included only for illustrative purposes.) The model could be improved by removing these two indicators and re-estimating coefficients for those indicators remaining in the model. When weak variables are removed from the model, the outcome is often that the coefficients on the remaining variables increase in magnitude and the number of extracted components decline.

TABLE 8.4 SPSS Component Matrix containing additional variables

Component	1	2	3
Number of days in past 7 days wheat was served in meal	.698	-.293	-9.075E-02
Number of days in past 7 days rice was served in meal	.476	-.435	-6.430E-02
Number of days in past 7 days meats were served in meal	.720	-.151	-8.171E-02
Number of days in past 7 days inferior food was served in meal	-.256	.232	.643
Quality of dwelling walls	.406	8.688E-02	.393
Household electricity	.563	-6.438E-02	.280
Per person expenditure on clothes and footwear	.629	4.306E-02	.153
Per person value of total assets	.454	.510	-.300
Percent of adults who completed secondary school	.565	-9.476E-02	.228
Value of appliances and electronics	.660	.357	1.027E-02
Value of radios owned by household	.565	.237	-3.203E-03
Days with enough to eat in past month	.402	-.464	-.353
Quality of latrine	.592	.196	.203
Number of bulls and cows	.142	.533	-.508

Extraction Method: Principal Component Analysis.
a 3 components extracted.

Even the most experienced analysts will run numerous combinations of variables to determine the combination of indicator variables that is most appropriate for explaining the underlying poverty component. The analysis of results can be repeated with alterations until the resulting model appears to be the most appropriate for the survey data. Ideally, the final version will capture more than one dimension of poverty (for example, food security, human resources, and asset accumulation), and no single group of measures will constitute the entire measure. Also, the final model will likely include no more than 20 indicators.

8.3.3 Step three: Revising the model until results meet performance requirements

PCA does not provide an easy way to come up with a best-fit model for determining a poverty index. The approach requires trial and error and a continual scrutiny of variables to determine which combination brings the most logical results. The primary strategy is to systematically screen the list of variables to use in the model without compromising the explanatory power of the poverty index.

The starting point for this is the component matrix, as was just described. In addition to the component matrix, several other techniques can be used to determine how the PCA poverty index model can be improved.

Level of explained common variance of the poverty index component

The SPSS output table "Explained Common Variance" displays the Eigen values calculated for each component. In the table, an Eigen value is shown for each of the calculated components. This size of Eigen value indicates the amount of variance explained by each component in the PCA model. The larger the Eigen value for a component, the more that component is 'explained' by the model's indicators. Table 8.5 is an example of the SPSS output table showing Eigen values. The second column from the left shows the calculated Eigen values for each component. The third column in the table shows the percent share of total variance each component explains.

If the model has been carefully screened to include only indicators of poverty, the first component is likely to explain variance associated with poverty. Eigen values and the associated level of variance explained by this component can guide in refining the model as variables are systematically added or deleted. As variables are deleted, the Eigen value for the poverty index component will change, as will the percent of common variance it explains. The change in share of explained variance can signal whether the addition or elimination of a variable improved or reduced the poverty index's explanatory power.

TABLE 8.5 Explained common variance

	Initial Eigen values			Sum of squared factor loadings for extraction		
Component	Total	Percent of variance	Cumulative percent	Total	Percent of variance	Cumulative percent
1	4.128	37.530	37.530	4.128	37.530	37.530
2	1.395	12.681	50.210	1.395	12.681	50.210
3	0.891	8.101	58.311			
4	0.830	7.541	65.853			
5	0.715	6.505	72.357			
6	0.704	6.396	78.753			
7	0.614	5.581	84.334			
8	0.505	4.588	88.923			
9	0.457	4.152	93.075			
10	0.406	3.687	96.762			
11	0.356	3.238	100.000			

Extraction method: principal components.

As a rule, an Eigen value of at least 1 is considered a minimum measure to regard the component as representing a common underlying dimension. In the table, only the first two components indicate that common variance is being measured. The first component, in this case the poverty index, explains 37.5 percent of the total variance; and the second, 12.7 percent. In general, because the model has been refined to create a measure of relative poverty, it is reasonable to expect the component explaining the most variance to be the poverty indicator.

Relative size of the communalities

Another means of testing the appropriateness of the model is to note in the model results the relative size of the communalities. Communalities represent the strength of the linear association among the variables and the components. Statistically they represent the same measure as R-squared in a regression analysis. The values of communalities range between 0 and 1, with higher numbers indicating a greater share of common variance explained by the extracted components.

Communalities indicate how well the indicators combine to identify different components. Since we are interested in only one of several shared components, communalities alone do not indicate the appropriateness of the variable for our model. Improving the measures for communalities will not improve the poverty index component if the added variables correlate strongly with components other than poverty.

Some variables may contribute to the explanatory power of a poverty factor but not account for variances captured by other common factors. As a result, variables may have low communality coefficients but still be relevant indicators for creating the poverty component. However, in general, communalities close to 0 (less than 0.1) signal that the variable in question may be a candidate for exclusion in subsequent runs. Table 8.6 is an example of the communalities table.

TABLE 8.6 Communalities table

Variable	Initial	Extraction
Number of days in past 7 days wheat was eaten	1.000	0.543
Number of days in past 7 days wheat was eaten	1.000	0.543
Number of days in past 7 days meat was eaten	1.000	0.597
Number of days in past 7 days meat was eaten	1.000	0.597
Extra monthly income spent on food	1.000	0.524
Extra monthly income spent on food	1.000	0.524
Members had enough to eat in past month	1.000	0.377
Members had enough to eat in past month	1.000	0.377
Household source of cooking fuel	1.000	0.461
Household source of cooking fuel	1.000	0.461
Household electricity use	1.000	0.506
Household electricity use	1.000	0.506
Aggregate value of all appliances and electronics	1.000	0.443
Aggregate value of all appliances and electronics	1.000	0.443
Value of radios	1.000	0.415
Value of radios	1.000	0.415
Percent of adults who completed secondary school	1.000	0.381
Percent of adults who completed secondary school	1.000	0.381
Percent of adults who can write	1.000	0.652
Percent of adults who can write	1.000	0.652
Per person expenditure on clothes and shoes	1.000	0.713
Per person expenditure on clothes and shoes	1.000	0.713
Family size	1.000	2.82E-02

Family size	1.000	2.82E-02
Materials used in house walls	1.000	0.198
Extraction method: principal component analysis		

The table shows that communalities ranging in value from 0.198 to a high of 0.652 can be considered within an acceptable range and all prove highly explanatory in creating the poverty component shown in Table 8.3. Only the indicator for family size has a communality coefficient near to zero.

Kaiser-Meyer-Olkin measure of sampling adequacy

The Kaiser-Meyer-Olkin (KMO) test is an index for comparing the magnitudes of the observed correlation coefficients with the magnitudes of the partial correlation coefficients. The smaller the value of the index, the less appropriate the model. In general, scores above 0.60 are acceptable, above 0.70 are good, above 0.80 are commendable, and above 0.9 are exceptional.

The first SPSS output table shows the results of both the KMO test and the Bartlett test of sphericity. If the table does not appear, then the model may contain variables that duplicate the same information. Check the variable list to see that there is no duplication. Table 8.7 is an example of how the table appears in the output. The table shows in the first cell of the right column the measured KMO for the model. The number, 0.855 in this case, is within the acceptable range for a well-specified model. The chi-square test is not used in this methodology because with sample sizes as large as 500, the test will almost always show less than .001 significance.

TABLE 8.7 KMO-Bartlett test

Kaiser-Meyer-Olkin measure of sampling adequacy	0.855
Bartlett test of sphericity	
Approximate chi-square	918.033
df	55
Significance of Bartlett	0.000

8.3.4 Step four: Saving component scores as a poverty index variable

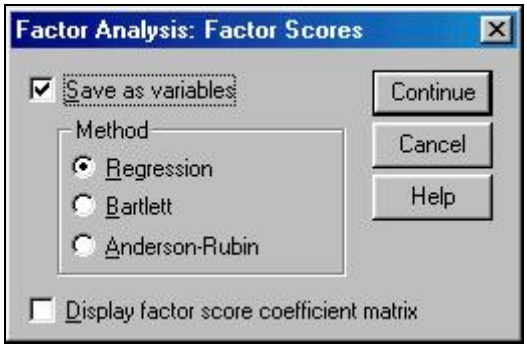
Once the final model for computing the poverty index is decided on, the sample size used to calculate the poverty component can be increased from the 300 nonclients to the full 500 households. This can be done within the FACTOR ANALYSIS dialogue box by removing the MFI status variable from the "Selection Variable" box. Using the full 500 sample size, rerun the model to register the final component calculations.

Run the final PCA model using the full 500 sample size and verify that no unusual results occur. If the measures of good fit decline slightly, do not re-specify the model. Because the random sample of MFI clients cannot be considered an unbiased representation of the local population, the MFI client cases are not used to set model specifications.

Using the final version of the PCA model, save the standardized values of the poverty component as a variable in the household data file. This is easily done from the FACTOR ANALYSIS dialogue box.

First, click on “Scores,” near the bottom of the main FACTOR ANALYSIS dialogue box, to open the FACTOR ANALYSIS: FACTOR SCORES dialogue box (Figure 8.7). In this dialogue box, check the box “Save as variables” and the circle “Regression” under “Method.” Hit “Continue. Second, open the FACTOR ANALYSIS: EXTRACTION dialogue box by clicking on “Extraction” in the FACTOR ANALYSIS dialogue box. Check the box “Number of factors” (see figure 8.6). This will cause the box to the right to be highlighted. Enter 1 to indicate that only the first component is to be saved as a variable. Rerun the PCA model. Check to ensure a new variable, “Factor regression score,” was created in the household file. Change the variable name to POVINDEX and add a variable definition such as ‘household poverty index’.

FIGURE 8.7 SPSS dialogue box to save component scores as variable



8.4 Properties of the poverty index variable created

The poverty index created through principal component extraction is estimated from standardized indicator values. This standardization is automatically made by SPSS before running the PCA. The poverty index is also in standardized form. Standardizing a variable works to strip away the units in which a variable is measured. A standardized variable has a mean of zero and a standard deviation of 1. Figure 8.8 shows the distribution of a poverty index in standardized form. In the graph shown, the poverty scores range from –2.51 to 3.72. Approximately two-thirds of households fall in the range between -1 and 1.

FIGURE 8.8 Histogram of the standardized poverty index

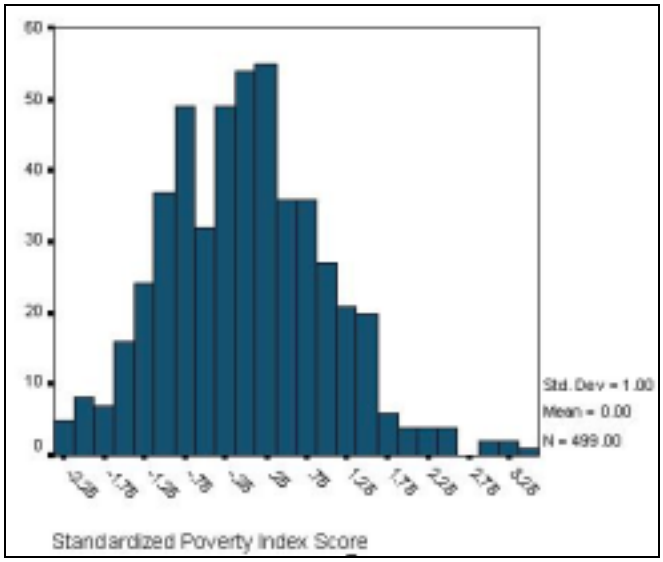
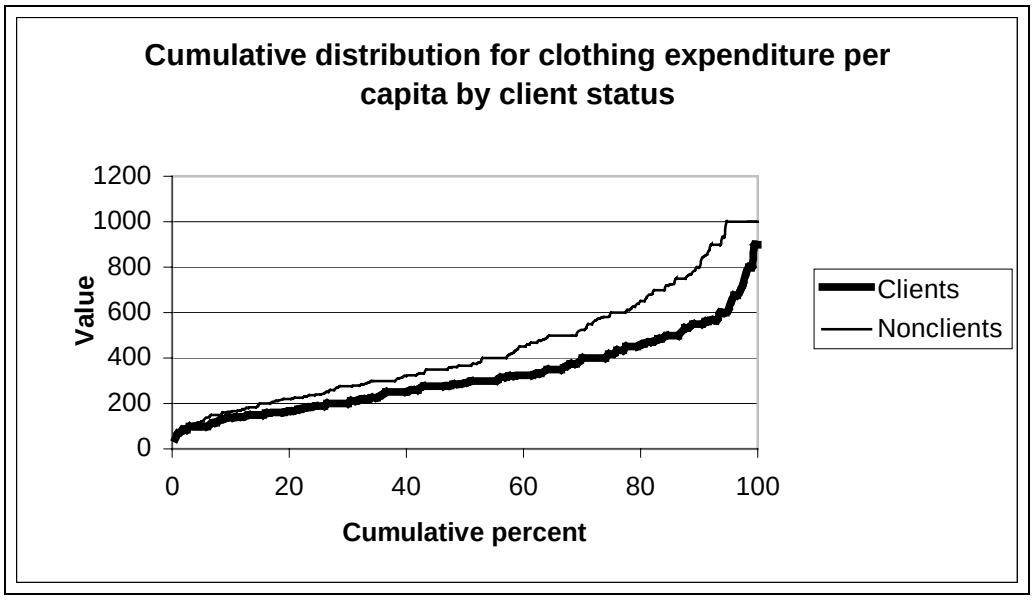


Figure 8.9 also shows the cumulative frequency of a different poverty index graphed for clients and nonclients. As the figure indicates, a fairly large margin of difference between the two groups exists except for the poorest of households, where differences between client and nonclient scores converge. For the poorest 10 percent of households, no difference is seen between client and nonclient poverty levels. However, for all other levels of relative poverty, clients appear poorer than nonclients. This can be cross-checked with the average poverty index score for clients against

nonclients. In our example, the average nonclient score is 0.22 and the average nonclient score is -.13, suggesting that on average clients are assessed as poorer than nonclients in the same area.

FIGURE 8.9 Cumulative frequency of client and nonclient scores



8.5 Using relative poverty terciles to interpret the poverty index

8.5.1 Defining the poor within the local population

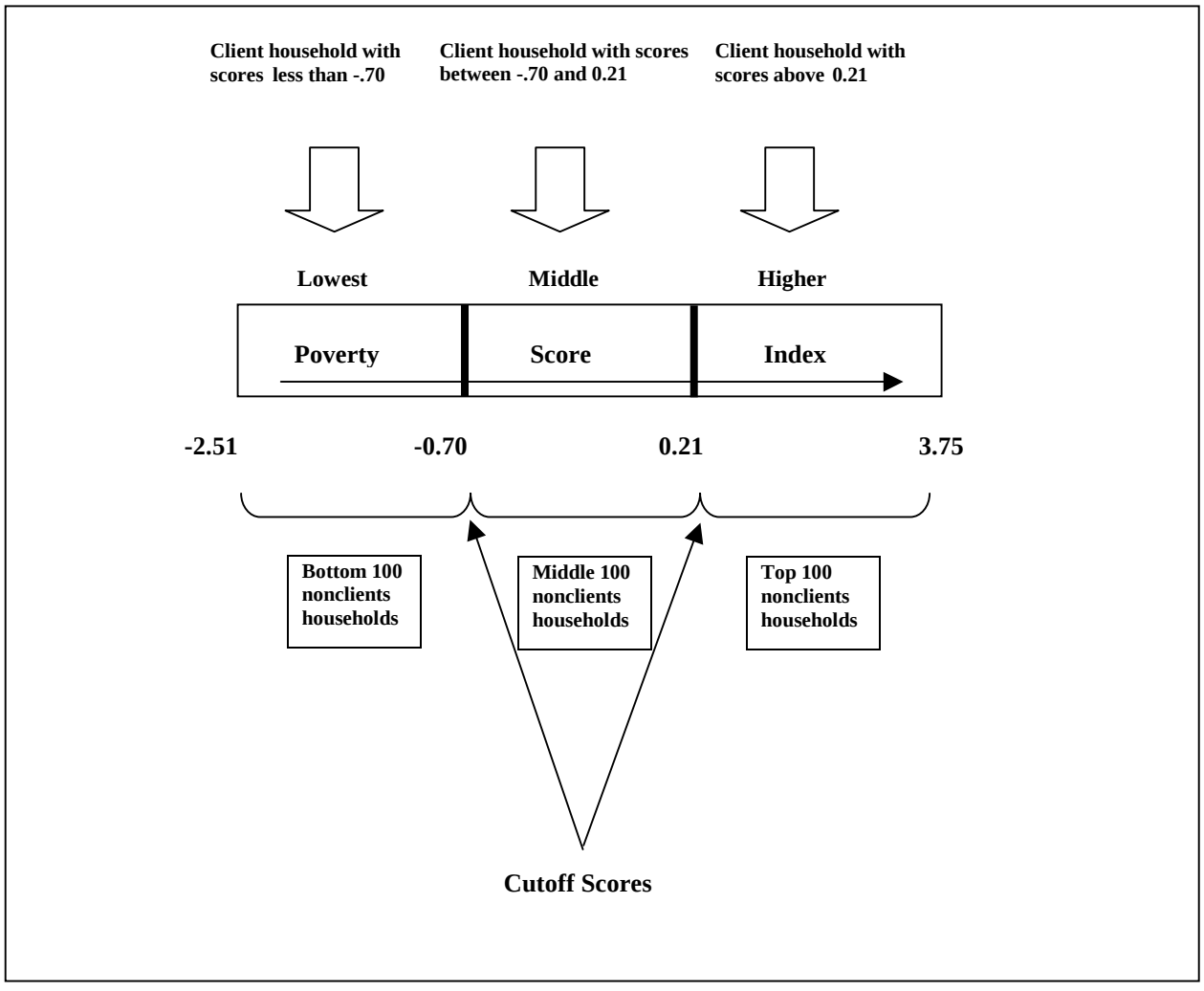
The creation of the poverty index assigns to each household a poverty ranking score. The lower the score, the poorer the household relative all others with higher scores. The scores of MFI client households and nonclient households can now be compared to indicate the extent to which the MFI reaches the poor. First, however, the share of the local population that is likely to fit the assessment's definition of poor must be decided on. If the researcher is interested in measuring the extent to which the MFI succeeds in reaching the poorest of the local population, an appropriate definition may be the poorest 20 percent of the local population. A broader definition of poor may assume that the lower half of the local population can be considered poor.

In this assessment methodology, a cutoff of 33 percent is used to define the poorest group within the local population. This decision is based on the usefulness of categorizing local populations into terciles that can be broadly interpreted to represent the lowest, middle and higher ranked groups of households ranked by relative poverty. The methodology can be adapted to include additional categorization, but other divisions should be reported in addition to the tercile results.

Each assessment study includes a random sample of 300 nonclient households and 200 client households. To use the poverty index for making comparisons, the nonclient sample is first sorted in an ascending order according to its index score. Once sorted, nonclient households are divided in terciles based on their poverty index score: the top third of the nonclient households are grouped in the "higher" ranked group, followed by the "middle" ranked group and finally the bottom third in the

"lowest" ranked group. Since there are 300 nonclients each group contains 100 households each. The cutoff scores for each tercile defines the limits of each poverty group. Client households are then categorized into the three groups based on their household scores. Figure 8.10 illustrates the use of cutoff scores to create poverty terciles from nonclient households. The cutoff scores of -.70 and +.21 were calculated from an actual case study example. Each poverty assessment will use different cutoff scores to group households. The steps involved in determining and applying these scores is described in the next section.

Figure 8.10 Constructing poverty groups



8.5.2 SPSS procedures for creating poverty terciles

Step 1: Limit sample to nonclients

Households are grouped into terciles first by using only the 300 nonclient households. From the **Data** menu, click on **Select Cases**, then **Select If** and use the displayed dialogue box to filter cases for only nonclients.

Step 2: Rank nonclient households to create three relative poverty terciles

Terciles of the poverty index are created by selecting **Rank Cases** from the **Transform** menu. In the dialogue box that opens, click on RANK TYPES and then click in the "Ntiles" box "under Options" and type in the number 3. This is shown in Figure 8.11. This will segment the sample of nonclients of 300 into three groups. If done correctly, approximately 33 percent of all nonclient households, or roughly 100 households, will be assigned to one of the three groups. To verify that the ranking was done correctly, run a frequencies test on the ranking variable automatically created in SPSS (NPOVINDE). This variable will begin with the letter "N" and add the first seven characters of the variable name for the poverty index. The SPSS dialogue boxes for ranking data are shown in figures 8.11 and 8.12.

FIGURE 8.11 SPSS dialogue box for ranking cases

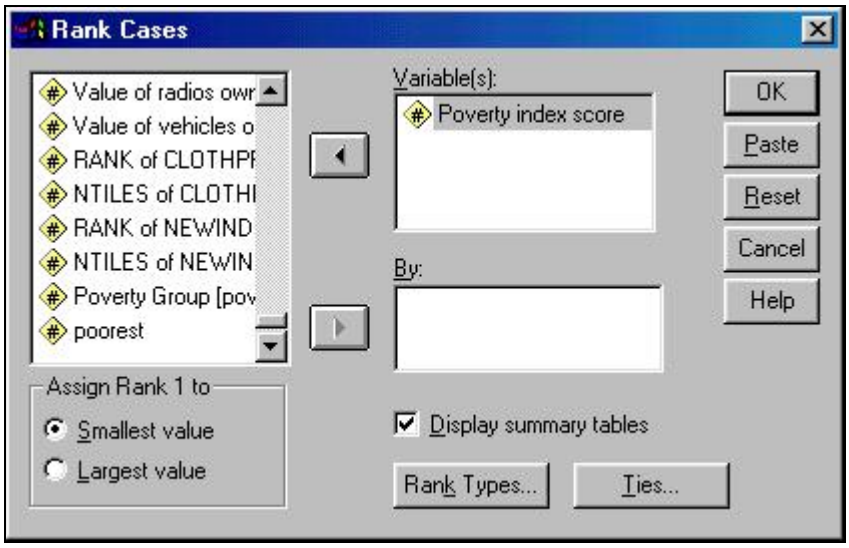
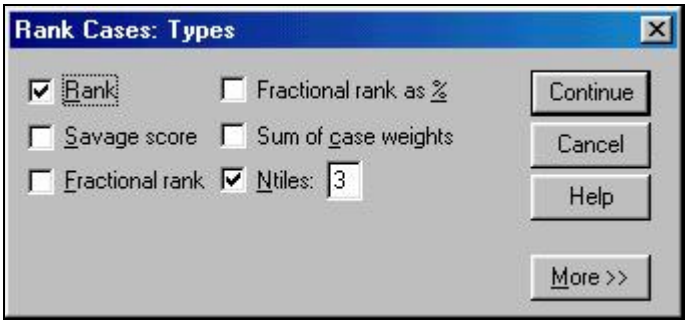


FIGURE 8.12 SPSS dialogue box for ranking data in cases



Step 3: Integrate MFI client households into relative poverty groupings

Each tercile created for nonclient households contains distinct value ranges of the poverty index. The maximum and minimum values for each range can be used to determine the appropriate categories to assign the MFI client households.

To create poverty terciles, first select only cases that are currently assigned to the middle poverty tercile. Click on **Select Cases** under the **Data** menu, then filter cases to only those where the poverty tercile equals 2.

From the **Analyze** menu, select the DESCRIPTIVE ANALYSIS option, then DESCRIPTIVES. In the dialogue box that opens, move the poverty index variable into the "Variables" box and click on "OK." The resulting table will look somewhat like that table 8.8. Note the minimum and maximum values, as these will be used to set boundaries for assigning the MFI client sample to the three terciles.

TABLE 8.8 Descriptive statistics of the middle tercile

	N	Minimum	Maximum	Mean	Std. deviation
Poverty index scores	100	–0.70134	0.21338	–0.2470421	0.2703259
Poverty index scores	100	–0.70134	0.21338	–0.2470421	0.2703259
Valid N (listwise)	100				
Valid N (listwise)	100				

Once the range of values for each tercile of nonclients is recorded, assign MFI client households according to the range in which each household's poverty index score falls. This is best done by computing a new variable, POVGROUP, that will list the group numbers for the full 500 households. Before computing, however, verify that all 500 households have values for the poverty index. If not, limit the sample to exclude all cases of missing values before proceeding with the following procedure.

Assign values to the variable POVGROUP as follows:

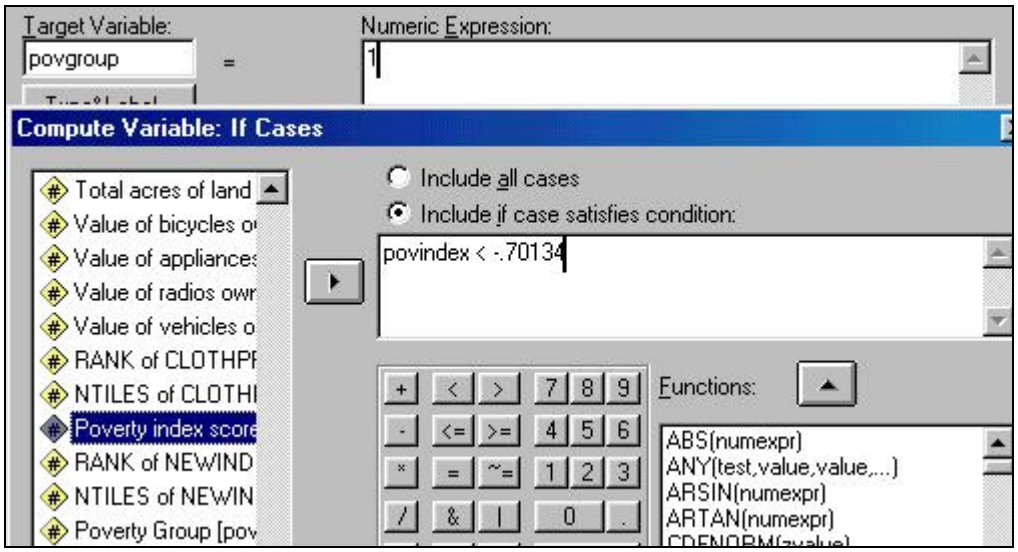
- 1—for all cases where the poverty index score is below the minimum value shown in your assessment's version of table 8.8 (-0.70134 in the example).
- 2—for all cases where the poverty index score is on or between the minimum and maximum values for your assessment's version of table 8.8 (-0.70134, 0.21338 in the example).
- 3—for all poverty index scores above the maximum value for your assessment's version of table 8.8 (0.21338 in the example).

Begin by selecting **Compute** from the **Transform** menu. In the dialogue box that opens, type in the new variable's name, POVGROUP, in the top left box, and in the box to the right enter "2." Click on "Continue," then "OK." This will result in all values of the new variable being 2.

Now, revise the newly computed variable by repeating the above process but this time, type in "1" in the top right text box and click on "If" to open the COMPUTE: IF dialogue box and set the "if" condition for POVGROUP = 1 to only those cases where the value of the poverty index variable is less than the minimum value of the poverty index shown in the descriptive table. In the example table shown above, this value was –0.70134. The values of POVGROUP should now show values of 1 wherever the poverty index value is below –0.70134. Figure 8.13 shows the SPSS dialogue boxes used to assign values for the POVGROUP variable.

Compute the POVGROUP values a final time to compute a value 3 where the condition holds for only poverty index values above the maximum level of the middle tercile. In our example, this amount is 0.21338.

Figure 8.13 SPSS dialogue boxes for assigning tercile values to client households



Step 4: Verify that all households have been categorized correctly. Once all cases have been assigned to a poverty group, run a frequencies table of POVGROUP to verify that the results are correct. Also verify that all cases with missing values for the poverty index also have missing values within POVGROUP. Use the SELECT CASES dialogue box (accessed from the **Data** menu) to set the "if" condition to "MISSING(povindex)." If these cases are found to have values of 1, recode the cases to SYSMIS using the **Recode** option in the **Data** menu. Finally, add a variable label, household ranking of relative poverty, and value labels for the new variable where 1 = lowest ; 2 = middle; and 3 = highest.

8.6 Assessing MFI poverty outreach by poverty groupings

Now that all cases for MFI clients and nonclients have been assigned to poverty groupings, comparing differences between the two distributions is possible. If the pattern of client households' poverty matches that of the nonclient households, client households would divide equally among the three poverty groupings just as the nonclient households, with 33 percent falling in each group. Hence any deviation from this equal proportion signals a difference between the client and the nonclient population. For instance, if 60 percent of the client households fall into the first tercile or lowest poverty category, the MFI reaches a disproportionate number of very poor clients relative to the general population.

Figure 8.14 shows the results of a recent case study highlighting significant differences in the poverty distribution between clients and nonclients. The graph shows that clients are overrepresented within the lowest tercile and underrepresented in the highest tercile. This would indicate that the

MFI is reaching a larger share of poorest households than what is found in the population in general. In contrast, the results of another case study found the opposite pattern. In Figure 8.15, the results indicate that MFI clients are underrepresented in the lowest tercile and overrepresented in the highest tercile.

To create graphs in SPSS similar to those shown in Figures 8.14 and 8.15, click on **Graphs**, then **Bar…** This will open the BAR CHARTS menu. Choose the "Clustered" option, and select "Summaries for groups of cases". Click the **Define** option to open the DEFINE CLUSTERED BAR dialogue box. In the top part of the box, select the option to show the " percent of cases". Then select the variable POVGROUP as category axis, and the client status variable to define clusters. Click on the **Titles…** option in the lower right corner to add titles for the graph.

FIGURE 8.14 Percent breakdown by poverty tercile, MFI clients and nonclients, results showing extensive poverty outreach

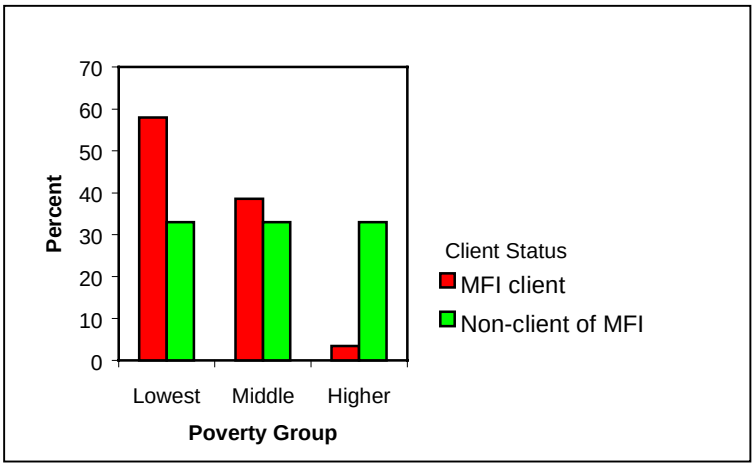
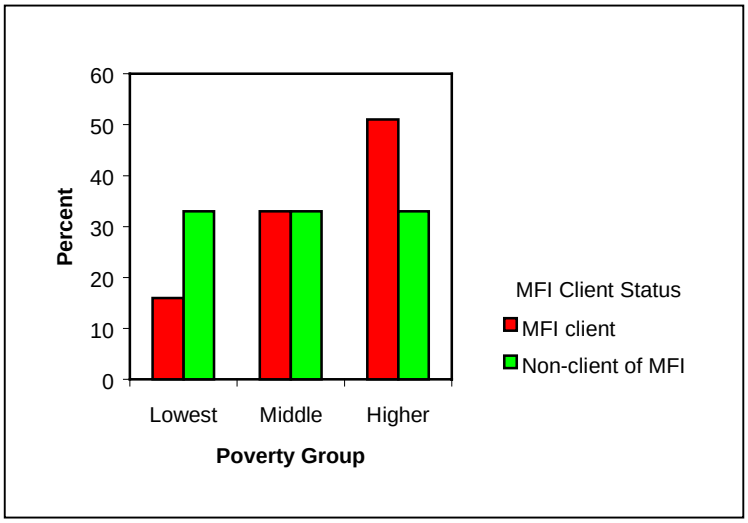


FIGURE 8.15 Percent breakdown by poverty tercile, MFI clients and nonclients, results showing lower poverty outreach



PART III

APPLYING THE RESULTS

CHAPTER NINE:
APPLYING THE RESULTS

A comprehensive assessment of an MFI must include an evaluation of how its poverty outreach record reconciles with its mission and program objectives. MFIs differ in terms of geography, their stated mission, the type of market niche they seek, their preference for a specific type of institutional culture, and a host of other factors. Ignoring these considerations or providing incomplete information on institutional details fails to tell a complete story. Interpreting results of the poverty assessment within the MFI context adds depth of understanding as to why the quantitative measurement of client and nonclient relative poverty differences turn out as they do.

This chapter will further guide users in applying the poverty index to make comparisons across programs and countries by explaining how summary ratios can be developed. These ratios can be used in conjunction with additional area-level and national-level information to interpret and compare poverty outreach of different MFIs. The final section of chapter will relate the entire assessment process to the MFI context and guide users in preparing a summary report of findings.

9.1 Comparing results at the local, area, and national levels

MFI outreach to the poor can be assessed at three levels:

- Local—the extent to which the MFI provides services to households at different poverty levels in surveyed area,
- Area—the extent to which the surveyed areas represent relatively poor parts of the country,
- National—the extent to which the country can be assessed as poor relative to all other countries.

While the first level of assessment has formed the core of this manual, an overall conclusion must explicitly account for area- and national-level considerations. An overall picture that takes into account all three levels can be developed and presented in a table to form the basis for making final comparisons. This added information then can be combined with qualitative analysis of the MFI's institutional context to interpret why the MFI's poverty outreach profile performs as it does.

9.1.1 Comparisons at the local level

If the pattern of poverty of client households were similar to those of the nonclient households, we would have expected client households to be distributed among the three poverty groupings in the same fashion as the nonclient households: 33 percent falling in each group. Hence any deviation away from this proportion signals a difference between the client and the nonclient population. Two ratios based on this deviation can be computed:

- Ratio 1: This is computed by dividing the percentage of client households that belong to the poorest group by 33, which is (by construction) the percentage of nonclient households who belong to this group. Hence the ratio reflects the extent to which the poorest households are represented in the client population. A ratio of 1 indicates that the proportion of the poorest households among the MFI's client is the same as in the general population. Ratios greater than 1 imply that that proportion of the poorest households among the MFI's clients is greater than that in the general population. On the other hand, ratios less than 1 imply that the proportion of the poorest households among the MFI's clients is less than in the general population.
- Ratio 2: This is computed by dividing the percentage of client households who belong to the less-poor group by 33. The ratio reflects the extent to which the less-poor households are represented in the client population. A ratio of 1 or less than 1 indicates that, compared with the nonclient population, a lesser proportion of client households falls into the less-poor group.

Ratio 1
--Percentage of clients belonging to lowest ranked poverty tercile divided by 33 percent --Higher values show more extensive outreach to the poorest households in the local area.
Ratio 2
--Percentage of clients belonging to the highest ranked poverty tercile divided by 33 percent --Higher values show more outreach to the better-off in local area.

9.1.2 Comparisons at the area level

As indicated in chapter 1, a local-level assessment of the relative poverty of MFI clients will not provide a complete picture if MFIs tend to locate in better-off or worse-off areas within a given country. In wealthier regions, the relatively poorer clients may still be better off, on average, than households living outside the operational area of the MFI, or conversely in poorer regions, higher ranked households may be worse off, on average, than households living outside the MFI's operational area.

Making assessments at the area or national level requires sampling of households outside the operational area of the MFI, escalating the cost of assessment exponentially and rendering it impractical. There are two options available for assessing the level of poverty in the MFI's operational area compared to other parts of the country. The first option is to collect information on area-level poverty measurements from various published sources. However, this is feasible only in countries where secondary information that is regionally disaggregated is available. In only a handful of developing countries are secondary data likely to be available that are disaggregated enough to allow comparisons between the operational area of the MFI and the rest of the country.

Further, when data are compiled from more than one source (which is likely), differences can exist in the division of areas used, the units of measure, definitions in terms, and the year in which the data were collected. Only if a standard methodology for measuring poverty is available which is countrywide and area-specific can be used, is a quantitative approach feasible. In all other cases, a second option can be applied.

The second option involves a qualitative assessment using an expert panel to rate the poverty level of the MFI operational area against the national standard. The steps involved in doing this are described below.

Step 1. Define areas to be assessed. Area level assessment should include the entire operational area of the MFI, not just the branches selected for the household survey. Make a list of all regions or branches and then within each list the names of localities where clients are located. When locations cannot be easily "recognized" by potential panel of experts (see step 2 below), each location should be mapped to the closest commonly understood set of geographic coordinates such as village names or local administrative units about which information can be relatively easily solicited. The list prepared should be arranged as in Table 9.1 with columns (1) and (2) filled in by the analyst.

TABLE 9.1 Worksheet for eliciting MFI area-level poverty ratings

MFI Operational Area (1)	Equivalent government administrative area(s) (2) TO BE FILLED BY ANALYST	Poverty level of general population (Score[a]) (3) **TO BE FILLED BY EXPERT**
MFI Region 1		
Within region 1 --Locality 1A --Locality 1B --Locality 1C --Locality 1D		
MFI Region 2		
Within region 2: --Locality 2A --Locality 2B --Locality 2C --Locality 2D		
MFI Region 3		
Within region 3: --Locality 3A --Locality 3B --Locality 3C --Locality 3D		

a. Score:
1 = Operational area ranks considerably below national average.
2 = Operational area ranks somewhat below national average.
3 = Operational area is at or around the national average.
4 = Operational area ranks somewhat above the national average.
5 = Operational area ranks considerably above national average.

Step 2. Identify the panel of experts. Key respondents for this assessment should be selected from a range of institutions, including major social-science research institutes, governmental or nongovernmental organizations involved in poverty alleviation programs, and well-known but independent poverty experts. It is extremely important to make sure that the panel of experts has direct knowledge on the operational area of MFI so that it is in a position to rank it against the

national standard. Officials in the ministries of health and agriculture and in the local government often have extensive knowledge about very specific locations within the country and are in a position to make comparisons across regions and different administrative boundaries. Select 8 to 10 experts who have different institutional backgrounds (government, social-science research institutes, and nongovernmental organizations involved in poverty alleviation programs).

Step 3. Set criteria for assessing area-level poverty. The area-based poverty assessment uses the opinions of a panel of experts to rate the overall poverty level in specific MFI operational areas against national average poverty levels. Panel members are asked to assign a score to each locality using the criteria below:

- 1 = Operational area ranks considerably below the national average.
- 2 = Operational area ranks somewhat below the national average.
- 3 = Operational area ranks at or around the national average.
- 4 = Operational area ranks somewhat above the national average.
- 5 = Operational area ranks considerably above the national average.

Only five levels are used to ensure that ranking is unambiguous and simple for the panel of experts to use. It is important that adequate guidance be provided to the expert panel to assist in ranking and scoring. It should be made clear that assessment should be based on due consideration of factors such as

- Wage and employment levels,
- Physical and social infrastructure,
- Literacy levels and
- Agricultural conditions (if rural).

Step 4. Elicit information from panel of experts. The worksheet shown in Table 9.1 is distributed to all members of the expert panel who are asked to fill in column 3 providing ranking for each locality within each region. If experts are unfamiliar with a particular locality, they should be asked to leave the particular cell blank rather than record a guess.

Step 5. Triangulation. Triangulation of information received from the panel of experts will be necessary, especially in cases where widely divergent views exist among the experts. This situation will arise, for example, when one expert assigns a score of five and another a score of one to the same locality. In such cases, the results of the survey should first be clearly tabulated showing each panel member's scores for each locality. Further, whenever scores deviate by more than three points, those experts with divergent opinions should be asked to provide a brief written explanation supporting their conclusion. The tabulated results, along with the explanations, should then be recirculated to the panel to give them the opportunity to change their previous ranking in view of the overall results and explanations provided. If changes take place, the process should be repeated until no changes are made.

The best-case scenario is one where an overall consensus eventually emerges. If complete consensus does not emerge but individual scores do not deviate by more than two points, then all expert opinions are given equal weight and average scores are computed as in Step 6 below. If individual scores deviate by three or more points, it is highly likely that some of the members have incomplete information. Under such a situation, the analyst should independently evaluate all explanations provided for logical consistency and overall credibility, and must decide which of the divergent opinion/s to discard. All available secondary data should be used in doing so and reasons for discarding should clearly and transparently be written in the final evaluation report. Once a decision has been made, average scores may be computed from the individual scores retained.

Step 6. Calculate a weighted average rating for the MFI operational area. The worksheet in Table 9.2 describes the process for computing the overall rating for the MFI's operational area. For each locality, average ratings are computed by adding all the scores for that locality and dividing this total by the number of expert responses. This average is entered in column 3 of Table 9.2.The overall rating for the MFI operational area is then computed as the **weighted** average of all locality-specific ratings, using the locality's share of the total MFI client base as the weighting factor. In order to do this, the number of active clients should be entered for each locality in column (4) and the client share of each locality in column (5). The client share is obtained by dividing the number of clients in the locality by the total client base of the MFI. The next step is to multiply columns (3) and (5) and place the result in column (6). The sum of this column is the weighted average poverty rating for the entire MFI area. It is suggested that the actual tabulation of weighted averages be done using a spreadsheet file such as Microsoft Excel.

TABLE 9.2 Worksheet for calculating MFI area-level poverty rating

MFI Operational Area (1)	Equivalent government administrative area(s) (2)	Poverty level of general population (3)	Number of active clients (4)	MFI weight based on share of client base (5)	Weighted poverty level (column 3 x column 5) (6)
Within region 1 --Locality 1A --Locality 1B --Locality 1C --Locality 1D					
Within region 2: --Locality 2A --Locality 2B --Locality 2C --Locality 2D					
Within region 3: --Locality 3A --Locality 3B --Locality 3C --Locality 3D					
Entire MFI operational area					

The result of the area-based assessment may be summarized as a ratio where the weighted MFI operational rating is divided by 3, the value assigned for the national average standard of living. Ratio values less than 1 indicate that the MFI operational area is predominantly below the national average in standard of livings for the general population, and values greater than 1 indicate the opposite case.

Ratio 3

--Indicates whether the MFI reaches poorer regions within the country.

--Values <1 indicate poorer regions being reached and >1 indicates less poor regions being reached.

9.1.3 Comparisons at the national level

Because the absolute level of poverty differs from country to country, intercountry comparisons cannot be made on the basis of intra-country poverty rankings, whether at the client or the area levels. For example, the relatively "poorest" households in Latin America could actually be better off than the relatively "less poor" households in Asia and Africa. For this reason, client- and area-level rankings have to be supplemented by national-level rankings if comparisons among countries have to be made.

National averages of real per capita incomes that take into account differences in prices of goods and services across countries are frequently used to rank countries by international agencies, such as the World Bank. However, in maintaining consistency with the indicator-based approach used throughout this manual, and recognizing the multidimensional nature of poverty, it is recommended that the Human Development Index computed by UNDP be used to make comparisons. HDI combines information on income with information on the achievements in health and education and can be used to make comparisons at the international level. The particular indicator that has been chosen is the ratio of the country's HDI to the average index of all developing countries together. Therefore, the higher the ratio, the better off the country, with ratios greater than 1 indicating that a country's poverty level is lower than the average and ratios less than 1 indicating that the country's poverty level is higher than average. Annex 4 lists the HDI for all countries and the average HDI for all developing countries.

Ratio 4

--Indicates whether the country of study is relatively poor compared to other developing countries.

--Values <1 indicate the country is poorer than the average for developing countries, and >1 indicates less poor than the average developing country.

Table 9.3 provides an example of three of the ratios recommended for comparing MFI poverty outreach across programs and countries, using data studies that were used to test and develop the assessment methodology contained in this manual—MFI C and MFI D.

The percentage of client households who are as poor as the poorest one-third of the nonclient population is much higher for MFI D (58 percent) than for MFI C (16 percent). The resulting Ratio 1 shows that while the relatively poor are strongly overrepresented among MFI D's clients (Ratio 1 is greater than one), they are underrepresented among MFI C's clients (Ratio 1 is less than one).

Similarly, the percentage of client households who are as well off as the least poor one-third of the nonclient population is higher for MFI C (51 percent) than for MFI D (3.5 percent). Ratio 2 shows that while the least poor households are overrepresented among MFI C's clients (Ratio 2 is greater than one), they are underrepresented among MFI D's clients (Ratio 2 is considerably less than one).

Ratio 3 was not tested in the initial case studies, due to a lack of reliable secondary data. This information would be crucial for comparing the relative poverty level of the MFI's operational area to the national average, thus placing the results of Ratios 1 and 2 in context. Further work is required to test the qualitative method outlined in Section 9.1.2 to fill this gap.

The absolute level of poverty differs from country to country, and this information must be taken into account to complete the poverty measurement exercise. For instance, it may be possible that the relatively "poorest" households in the operational area of MFI D are actually better off than the relatively "less poor" households in the operational area of MFI C, as implied in Ratio 4. The latest Human Development Report (2000) assigns an HDI of 0.563 for MFI D's country and 0.483 for MFI C's country. The HDI for all developing countries taken together is 0.542. Hence Ratio 4 indicates that standards of living are higher in MFI D's country than in MFI C's country.

TABLE 9.3 Relative poverty ranking of client vs. nonclients

Percentage/ Ratio	MFI C	MFI D
Percent of client households who are as poor as the poorest one-third of the nonclient population	16%	58%
Ratio 1	0.48	1.76
Percent of client households who are as well of as the least poor one-third of the nonclient population	51%	3.5%
Ratio 2	1.55	0.11
Ratio 4	0.89	1.04

9.2 Comparing results to the MFI's mission and objectives

A comprehensive assessment of an MFI's poverty outreach effort must be placed in the context of its mission and program objectives. MFI programs differ not only in terms of geography, but also in the mission they define for themselves, the type of market niche they seek for themselves, their preference for a specific type of institutional culture, and a host of conditionalities imposed by other external actors at various levels. In the case studies presented in Table 9.2, for instance, while membership in MFI C is share-based and open to all individuals, MFI D explicitly targets its services to the poorest households in its operational area. A poverty assessment exercise that ignores or provides incomplete information on these types of institutional details fails to tell a complete story and therefore can be easily put to potential misuse. It is recommended that the poverty assessment

exercise be used only in conjunction with an overall institutional appraisal using an instrument like CGAP's appraisal format and including contextual issues like:

1. The declared mission or vision guiding the MFI's goal and operational procedures,
2. The MFI's product and service range and its implication on overall client selection,
3. The specific geographical area focus of the MFI,
4. The assessment of conditionalities imposed by investors, promoters, governments, and local communities,
5. The nature of political and other external constraints faced by the MFI,
6. The state of market competition faced by the MFI,
7. The stage of institutional development of the MFI.

9.3 Reporting the findings

The results of the poverty assessment can best be presented in the form of a written report. The outline for the final report should include the following:

1. Executive summary of major findings,
2. Introduction and objectives of the assessment,
3. Background on the MFI—mission, structure (geographical and structural), performance in recent years, and future directions,
4. Summary of the methodology, including the sampling frame, the content of the questionnaire, implementation, and analysis of data,
5. Discussion of constraints and limitations of data and interpretation of results,
6. Results from creating the poverty index, including calculation of ratios, supported by figures and tables of statistical outcomes,
7. Qualitative discussion that interprets results, including any noteworthy differences between client and nonclient groups,
8. Annexes consisting of the statistical test results, the questionnaire, and other relevant materials.

ANNEX 1:
ASSESSING POVERTY THROUGH ALTERNATIVE APPROACHES

There are three principal methods for assessing the poverty level of a household:

- Household expenditure analysis and computation of a poverty line,
- Rapid appraisal or participatory appraisal methods,
- Indicator analysis, using an index of relative poverty.

These methods and the advantages and disadvantages of using each as practical tools are briefly described below as background information for the evaluator. A number of references are given for readers who wish to expand their knowledge on these methods.

Detailed household expenditure survey

The expenditure survey method is widely used in nationally representative households surveys, such as the Living Standard Measurement Survey, conducted by the World Bank. The standard practice in poverty analysis has been to use household total expenditure as the primary measure to evaluate the standard of living of households. It is argued that total expenditure expresses a good measure of the household's command over goods and services it chooses to consume.

The basic criteria used in assessing whether a household is poor or not is based on an evaluation of whether its income is sufficient to meet the food and other basic needs of all household members to lead a healthy and active life. To make the assessment, a basket of goods and services satisfying a pre-set level of basic needs is constructed. This basket corresponds to local consumption patterns, and is valued at local consumer prices to compute the minimum cost of acquiring this basic-needs basket. The value of the basket of minimum food, goods, and services is then called the "poverty line." This poverty line is most commonly expressed in per-capita terms. If the per-capita income of household members is below the poverty line, the household and its members are considered poor. If this does not hold, the household is categorized as nonpoor.[4]

The advantage of this method is that it is a widely accepted and fairly precise tool in measuring poverty, as far as the income dimension of poverty is concerned. The poverty line method allows for comparisons between clients and nonclients of MFIs within one area of a country and between countries. However, the data requirements of this method are very steep, and comprehensive standardized questionnaires are needed. The standard practice is to record food consumption data, using a recall period of one week, and using a combination of monthly or yearly recall periods to collect information on various nonfood items. Even though poor households in developing countries consume a small number of goods, given the long recall periods, accuracy in reporting is always a concern. A more accurate method is to require households to maintain a written diary of

[4] For further references on household expenditure surveys and the poverty line, see for example, Aho, Larivière, and Martin 1998; Chung et al. 1997; and Lipton and Ravaillon 1995).

expenditures, but this is hardly feasible in countries or environments where illiteracy is endemic. Second, even if consumption items could be accurately recalled, there are several other problems: ways have to be found to value home-produced foods when market information is lacking; irregular weights and measures make fixing quantities problematic; information on a number of high-value items, such as the rental value of housing, is likely to be seriously deficient. Given these difficulties, it is likely that collected data on household expenditures will be quite inaccurate. Of course, the scale of these problems can be substantially minimized by extensive training of interviewers, multiple household visits, and cataloging of informal weights and measures. However, the effect on survey cost and time to control for potential errors is likely to be exponential.

Moreover, the analysis of expenditure data necessitates advanced skills in statistical data analysis. This requirement translates into high costs for data collection as well for analysis. Another drawback of this method is that the definition of the minimum bundle of food and nonfood services required to achieve a minimum standard of living can be ambiguous in international comparisons if the minimum bundles of food and nonfood consumer items differ across countries.

The costs of the survey could potentially be reduced if the evaluator has access to benchmark data from a recently undertaken national household survey on poverty. If such data is accessible, the analyst may choose to undertake a similar household survey only for MFI clients, and to compare those results with the national benchmark (see, for example, Navajas et al. 2000). While this approach can reduce costs, it is only feasible in countries that have recently undertaken a national poverty study. However, in many developing countries, such data are either not available, outdated, or difficult and costly to access. In terms of costs of analysis, considerable time will need to be spent by the evaluator on getting familiar with the national data.

In summary, while the method can be considered as a reliable and valid assessment of poverty, it is far too costly, time-consuming, cumbersome, and analytically demanding to be chosen as the most practical method for assessing the poverty level of microfinance clients.

Rapid appraisal and participatory appraisal

Two other methods used for poverty assessment are Rapid Appraisal (RA) or Participatory Appraisal (PA). These methods are often thought to be the same since they seek input by the community and its members, using similar techniques—such as wealth ranking and community mapping. There are differences, however (Bergeron 1995). The ultimate goal of PA is empowerment of the target group, requiring extensive participation by the community and assumes an open research and development agenda. This can not be done quickly, that is, within one or two days. RA methods, on the other hand, are meant to provide evaluators with data on the community in a very short time. RA requires the participation of the community, but the time frame is short (usually a one-day visit to the community) and the agenda of the inquiry is predetermined.

RA and PA methods are widely used and accepted tools for identifying vulnerable groups in a community. They are extensively used by development programs and institutions, including (MFIs), for targeting services to poorer clients (Hatch and Frederick 1998). The RA method in particular has relatively low time requirements for data collection. While these methods can be well suited for targeting and for the participatory design of development projects and services, a number of

disadvantages exist for poverty assessment for purposes of regional, national, or international comparisons. First, the results are difficult to verify, as they stem from the subjective rating of the community members on who is poor in the community and who is not. Second, the approach is likely to find poor people in every community, and the percentages of poor people may not vary much across villages. In other words, the method may be consistent in finding the poorest third in one village, but it may not be consistent in finding in which communities reside the poorest third of an entire region. Finally, the PA method requires skillful and experienced communicators. For national and international comparisons, there could be concern about the bias that is introduced by the way that PA is implemented.

Indicator-based method

Another method is to identify a range of indicators that reflect powerfully on the different dimensions of poverty and for which credible information can be quickly and inexpensively obtained. Once information on a range of indicators has been collected, they may be aggregated into a single index of poverty. Desirable attributes of poverty indicators are reviewed in the next section.

One well-known application of this method is the Human Development Index (UNDP 1999). It is based on three components: educational attainment, life expectancy at birth, and per capita income adjusted for purchasing-power parity dollars. The latter two indicators are costly to measure in surveys, and therefore are not operational. Another example is the housing index, which is used by many MFIs (in particular in South and Southeast Asia) for targeting financial services to poorer clients. Its advantage is that the list of indicators feeding into the housing index, such as quality of roof or walls of the house, can be obtained very quickly by visual inspection of the house. A major disadvantage of this method is that it focuses only on one dimension of poverty (housing), while neglecting the other ones, such as food security and human resources. Further, the housing index may not be applicable when housing is homogeneous in the community, or when housing is not an important poverty dimension, such as in a region with a good climate. In principle, the time and cost requirements of the indicator method in terms of data collection and analysis can be relatively low if the number of indicators in the poverty index are limited. It can be valid if several dimensions of poverty are included. For these reasons, the indicator method was chosen as the one to be developed for the assessment of the poverty level of microfinance clients.

Overall, a good indicator is a measure that is easily observable, verifiable, and objectively describes poverty. A wide range of indicators is recommended to capture aspects of an underlying dimension of relative poverty within households.

Two main types of indicators can be used to assess the actual level of poverty of households: indicators on income and indicators on consumption. Studies comparing different indicators based on income and on consumption conclude that recommending one alternative measure over another is difficult (Skoufias, Davis and Soto 2000). However, consumption over time (seasons or years) is more stable than income, and households provide information more easily on what they consume than on what they earn. For this reason, this tool relies on selected indicators of consumption, although selected indicators expressing means available to the household to increase its standard of living are also included.

In coming up with reliable indicators, the key challenge is to identify key components of consumption that are either unambiguous measures of poverty in themselves (such as incidence of hunger) or those that correlate well with—or are good proxies of—total household expenditures. Hence, it is not necessary to compile all the food and nonfood expenditures of a household, since some types of expenses can be closely related to the level of poverty of a household, while others are not. Studies have shown that the proportion of clothing and footwear expenditures in the household budget remains stable at different income levels, around 5 to 10 percent of total expenses (Aho et al. 1998; Minten and Zeller 2000). A recent study by Morris, Carletto, Hoddinott, and Christiaensen (1999) found clothing expenditure as one of the expenditure components that increased proportionally with total household expenditures. Since clothing, unlike food commodities, usually requires a purchase of either the finished garment or materials to make a garment, it also avoids the valuation problem with food consumption or expenditure.

ANNEX 2:
LIST OF INDICATORS AND SCORING

Ranking of the indicators

A score has been attributed to each indicator according to the following guidelines:

M	Statistically determinant in some statistical **m**odels
N	**N**ationally valid (can be used in different local contexts, urban versus rural)
O	Not too sensitive a question, can be asked **o**penly
P	**P**ractical (can be observed as well as asked)
Q	High-**q**uality indicator (is sensitive in discriminating poverty levels)
R	**R**eliable (low risk of falsification or error; also possible to verify)
S	**S**imple (direct answer versus computed information)
T	**T**ime-efficient (can be answered rapidly)
U	**U**niversal (can be used in different countries)

When the indicator fulfills the requirement, it is marked by an upper case letter. When the indicator fails to fulfill the requirement, it is marked by a lower case letter. The score of an indicator is the total of the upper case, and it ranges from 0 to 9.

1 Means to achieve welfare

Human capital

FAMILY STRUCTURE

6	Number/age of adult (18-55), male/female	M N O p q R S t U
7	Number/age of preschooler (0-6)	M N O p q R S t U
6	Number/age of children (7-17), male/female	M N O p q R S t U
6	Number/age of old people (>55)	M N O p q R S t U
6	Female-headed household	M N O p q R S t U
6	Number of disabled persons	m N O p q R S T U
6	Number of women who had first child before 16	m N O p q R S T U

EDUCATION

5	Number of school-age children in school	m N O p q R S t U
6	Distance to school	m n O P q R S T U
7	Level of adult literacy	M N O p q R S T U
6	Years of schooling of adults	M N O p q R S t U
7	Head of household last grade completed	M N O p q R S T U

INCOME

Agricultural income

4	Average monthly/annual household income	M N o p Q r s t U
6	Source of agricultural income (food crop/cash crop/livestock)	M N O p q R S T u
6	Employment status (self-employed/tenant)	m N O p q R S T U
4	Last year's crop yield	m N O p q r s T U

Nonagricultural income

4	Average monthly/annual household income	M N o p Q r s t U
6	Employment status	m N O p q R S T U
6	Source of nonagricultural income	M N O p q R S T u
6	Number family/wage employees in microenterprise	m N O p q R S T U
7	Number of adult wage earners	m N O p Q R S T U
7	Number of adults unemployed	m N O p Q R S T U
4	Dowry/bride price level	m N O p q r S T u
7	Child Labor	m N O p Q R S T U

Transfers

6	Remittances from migrant member of family (national/abroad)	M N O p Q r S T u
5	Pensions (old age)	m N O p q r S T U
3	Gifts from family, friends, neighbors	m N O p q r s t U
6	Welfare pension	m N O p Q r S T U

LIABILITIES

5	Debts with financial institutions	M N o p q r S T U
4	Debts with informal moneylenders/shopkeeper	m N o p Q r S t U
3	Debts with friends and family	m N o p q r S t U

Assets

LAND

6	Landless (yes/no)	M n O p Q r S T U
5	Amount of land owned, leased	M n O p Q r s T U
5	Quality of the land	m n O p Q r S T U
3	Plot size	m n O p q r s T U
4	Market value of land	m n O p Q r s T U
5	Secure land tenure	m n O p q R S T U
6	Access to irrigation	M n O p Q R S T u
3	Size of irrigated/nonirrigated lands	m n O p Q r S t u
6	Use of agricultural inputs	m n O p Q R S T U

OTHER PRODUCTIVE ASSETS

7	Number/type/value of animal	M N O p Q R S t U
5	Number/type/value of trees	m N O p q r S T u
5	Number/type/value of buildings, machinery, equipment	m N O p Q R s t U
7	Number/type/value of car/motorcycle/bicycle	m N O p Q R S T U
6	Level of monetary savings in financial institution	m N O p Q r S T U
4	Level of loans given to friends, family, neighbors	m N O p q r S t U

NONPRODUCTIVE ASSETS

5	Number and type of cooking utensils	m N O p q R S t U
5	Number and type of jewelry	m N O p q R S t U
7	Electronic devices (radio, TV, phone)	m N O p Q R S T U

2 Basic needs

Health

5	Immunization of children	m n O p q R S T U
6	Number of working days lost to sickness	m N O p q R S T U
6	Incidence of contaminated water/food related disease	m N O p Q R S t U
5	Incidence of domestic hygiene related disease	m N o p Q R S t U
5	Number of children under 6 who have died of illness	m N o p q R S T U
6	Access to medical services	m N O p q R S T U
6	Children born in hospital/clinic or at home (birth attendant)	m N O p Q R S t U

Food/water

7	Number of meals a day	m N O p Q R S T U
6	Daily caloric intake	M N O p Q R s t U
8	Weekly intake of meat(/fish/luxury food)	M N O p Q R S T U
7	Staple substitution (cheaper staple)	m N O p Q R S T U
7	Need to buy staple food at lean season	m N O p Q R S T U
5	Level of malnourished children	m N O p Q R s t U
7	Type of access to potable water	m N O P q R S T U

Shelter

	Housing index	
7	1. Size of building	m N O P q R S T U
7	2. Number of stories	m N O P q R S T U
7	3. Structure condition	m N O P Q R S T u
7	4. Roof material	m N O P Q R S T u
7	5. Wall material	m N O P Q R S T u
6	6. Electricity supply	m N O P q R S T u
7	7. Piped water supply	m N O P Q R S T u
7	8. Vehicle	m N O p Q R S T U

5	Market value of house	m N O p Q r S t U
6	House ownership	m N O p q R S T U
6	Threat of eviction (urban)	m n O p Q R S T U
7	Type of roof/walls/floor	m N O P Q R S T u
6	Lighting source	m N O P q R S T u
6	Cooking fuel source	m N O p Q R S T u
7	Type of latrine	m N O p Q R S T U
5	Bathing/washing facilities	m N O p q R S T u
5	Sleeping conditions	m N O p q R S T u
4	Furniture	m N O p q R S t u

Expenses

5	Food expenses each week	M N O p Q r s t U
6	Purchasing staple food more than once a week	M n O p q R S T U
5	Nonfood expenses each week	M N O p Q r s t U
5	Share of food expenses in total budget	m N O p Q r s T U
5	Amount of unusual expenses during last month	m N O p q R s T U
6	Home rent/home purchase installment	m N O p q R S T U
5	Cost of recent home improvement	m N O p q R S t U

3 Other aspects of welfare

Security

6	Number of months without enough food	m n O p Q R S T U
6	Number of months of migration	m N O p q R S T U
5	Number of failed harvests in last three years	m n O p q R S T U
6	Number of natural disasters in last three years	m N O P q R S T u
4	Strategy in case of unexpected shock in income flow, open question	m N O p q R s t U

Social status

6	Caste membership	m N O p Q R S T u
5	Ethnic group	m N O p q R S T u
5	Minority group	m N O p q R S T u
4	Level of participation in local organization	m N O p q R s t U
4	Access to local elites	m N O p q R s t U
7	Gender	M N O p q R S T U
6	Marital status	m N O p q R S T U

Local environment

5	Distance to vehicle road	m n O P q R S T u
7	Distance to bank/post office/public transport/health services/school	M n O P q R S T U

ANNEX 3:
RECOMMENDED QUESTIONNAIRE

Assessing Living Standards of Households

International Food Policy Research Institute

A study sponsored by the Consultative Group to Assist the Poorest (CGAP)

Section A Household Identification

A1. Date (mm/dd/yyyy): __/__/____

A2. Division code: ☐☐

A3. MFI unit code: ☐☐

A4. Group code: ☐☐

A5.Group name: ☐

A6. Household code: ☐☐

A7. Household chosen as (1) client of MFI, or (2) nonclient of MFI? ☐

A8. Is household from replacement list? (0) No (1) Yes ☐

A9. If yes, the original household was (1) not found or (2) unwilling to answer, or (3) client status was wrongly classified: ☐

A10. Name of respondent:

Name of the household head: ☐

Address of the household: ☐

A11. Interviewer code: ☐☐ A12. Date checked by supervisor (mm/dd/yyyy): ___/___/____

A13. Supervisor signature: _____

Section B. Family Structure

B1. Adults members of household (aged 15 and above)

ID code	Name	Status of the head of the HH[a]	Relation to head of HH[b]	Sex[c]	Age	Max. level of school-ing[d]	Can write[e]	Main occupa-tion, current year[f]	Current mem-ber of study MFI[e]	Amount of loan borrowed from study MFI	Clothes/Foot-wear expenses for the last 12 months in local currency[g]
1	(HH head)										
2											
3											
4											
5											
6											
7											
8											

[a](1) single; (2) married, with the spouse permanently present in the household; (3) married with the spouse migrant; (4) widow or widower; (5) divorced or separated; (6) living mostly away from home but contributing regularly to household.
[b](1) head of the household; (2) spouse; (3) son or daughter; (4) father or mother; (5) grandchild; (6) grandparents; (7) other relative; (8) other nonrelative.
[c](1) male; (2) female.
[d](1) less than primary 6; (2) some primary; (3) completed primary 6; (4) attended technical school; (5) attended secondary; (6) completed secondary; (7) attended college or university.
[e](0) no; (1) yes.
[f](1) self-employed in agriculture; (2) self-employed in nonfarm enterprise; (3) student; (4) casual worker; (5) salaried worker; (6) domestic worker; (7) unemployed, looking for a job; (8) unwilling to work or retired; (9) not able to work (handicapped).
[g]In order to get an accurate recall the clothes and footwear expenses for each adult are preferably asked in the presence of the spouse of the head of the household. If the clothes were sewn at home, provide costs of all materials (thread, fabric, buttons, needles).

B2. Children members of household (from 0 to 14 years)

ID code	Name	Age	Clothes/ Footwear expenses for past 12 months, in local currency[a]
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

[a]Clothes and footwear expenses are asked for once those for adults have been recorded, and in the presence of the spouse of the head of the household. In case of ready-to-wear clothing and footwear items, include full price. In other cases, include cost of fabric, cloth as well as tailoring and stitching charges

Section C. Food-Related Indicators

(Both the head of the household and his or her spouse should be present when answering for this section.)

C1. Did any special event occur in the last two days (for example, family event, guests invited)?

(0) No (1) Yes ☐

C2. If no, how many meals were served to the household members *during the last 2 days*? ☐

C3. If yes, how many meals were served to the household members *during the 2 days preceding the special event*? ☐

C4. Were there any special events in the last *seven days* (for example, family event, guests invited)? (0) No (1) Yes ☐

(If "Yes," the "last seven days" in C5 and C6 should refer to the week preceding the special event.)

C5. During the *last seven days*, for how many *days* were the following foods served in a main meal eaten by the household?

Luxury food	**Number of days served**
Luxury food 1	
Luxury food 2	
Luxury food 3	

C6. During the *last seven days*, for how many *days* did a main meal consist of an inferior food only? ☐

C7. During the *last 30 days*, for how many days did your household not have enough to eat everyday? (0) No (1) Yes ☐

C 8. During the *last 12 months*, for how many months did your household have at least one day without enough to eat? (0) No (1) Yes ☐

C9. How often do you purchase the following?

Staple	**Frequency served**
Staple 1	
Staple 2	
Staple 3	

(1) Daily (2) Twice a week (3) Weekly (4) Fortnightly (5) Monthly (6) Less frequently than a month ☐

C10. For how many weeks do you have a stock of *local staples* in your house? ☐

C11. If your household earnings increased by (US$10–$20), how much of that would you spend on purchasing additional food? (Estimate amount as 5% of GDP per capita.) ☐☐☐

(*Note:* Does not include alcohol and tobacco.)

Section D. Dwelling-Related Indicators

(Information should be collected about the dwelling in which the family currently resides.)

D1. What is the ownership status of dwelling? (1) Built on squatter land (2) Owned (3) Given by relative or other to use (4) Provided by government (5) Rented ☐

D2. How many rooms does the dwelling have? (Include detached rooms in same compound if same household.) ☐☐

D3. What type of roofing material is used in main house? (1) Tarpaulin, plastic sheets, or branches and twigs (2) Grass (3) Stone or slate (4) Iron sheets (5) Brick tiles (6) concrete ☐

D4. What type of exterior walls does the dwelling have? (1) Tarpaulin, plastic sheets, or branches and twigs (2) Mud walls (3) Iron sheets (4) Timber (5) Brick or stone with mud (6) Brick or stone with cement plaster ☐

D5. What type of flooring does the dwelling have? (1) Dirt (2) Wood (3) Cement (4) Cement with additional covering ☐

D6. What is the observed structural condition of main dwelling? (1) Seriously dilapidated (2) Need for major repairs (3) Sound structure ☐

D7. What is the electricity supply? (1) No connection (2) Shared connection (3) Own connection ☐

D8. What type of cooking fuel source primarily is used? (1) Dung (2) Collected wood (3) Purchased wood or sawdust (4) Charcoal (5) Kerosene (6) Gas (7) Electricity ☐

D9. What is the source of drinking water? (1) Rainwater, dam, pond, lake or river (2) Spring (3) Public well—open (4) Public well—sealed with pump (5) Well in residence yard (6) Piped public water (7) Bore hole in residence ☐☐

D10. What type of toilet facility is available? (1) Bush, field, or no facility (2) Shared pit toilet (3) Own pit toilet (4) Shared, ventilated, improved pit latrine (5) Own improved latrine (6) Flush toilet, own or shared ☐

E. Other Asset-Based Indicators

E1. Area of land owned: Agricultural _____ Nonagricultural _____

Value of land owned: Agricultural _____ Nonagricultural _____

E2. Number and value of selected assets owned by household. (Ask household to identify any assets purchased with MFI loan and eliminate these from the table below.)

Asset type and code	Number owned	Resale value at current market price
Livestock		
1. Cattle and buffalo		
2. Adult sheep, goats, and pigs		
3.Adult poultry and rabbits		
4. Horses and donkeys		
Transportation		
5. Cars		
6. Motorcycles		
7. Bicycles		
8. Other vehicles		
9. Carts		
Appliances and electronics		
10. Televisions		
11. Video cassette recorders		
12. Refrigerators		
13. Electric or gas cookers		
14. Washing machines		
15. Radios		
16. Fans		

E3. What is your overall assessment of the general wealth levels of MFI clients? (1) Poor (2) Average (3) Rich (4) Don't know MFI ☐

ANNEX 4:
HUMAN DEVELOPMENT INDEX (HDI)*

Rank		Value (1998)
High Human Development		
1	Canada	0.935
2	Norway	0.934
3	United States	0.929
4	Australia	0.929
5	Iceland	0.927
6	Sweden	0.926
7	Belgium	0.925
8	Netherlands	0.925
9	Japan	0.924
10	United Kingdom	0.918
11	Finland	0.917
12	France	0.917
13	Switzerland	0.915
14	Germany	0.911
15	Denmark	0.911
16	Austria	0.908
17	Luxembourg	0.908
18	Ireland	0.907
19	Italy	0.903
20	New Zealand	0.903
21	Spain	0.899
22	Cyprus	0.886
23	Israel	0.883
24	Singapore	0.881
25	Greece	0.875
26	Hong Kong, China (SAR)	0.872
27	Malta	0.865
28	Portugal	0.864
29	Slovenia	0.861
30	Barbados	0.858
31	Korea, Rep. of	0.854
32	Brunei Darussalam	0.848
33	Bahamas	0.844
34	Czech Republic	0.843
35	Argentina	0.837
36	Kuwait	0.836
37	Antigua and Barbuda	0.833
38	Chile	0.826
39	Uruguay	0.825

Rank		Value (1998)
40	Slovakia	0.825
41	Bahrain	0.820
42	Qatar	0.819
43	Hungary	0.817
44	Poland	0.814
45	United Arab Emirates	0.810
46	Estonia	0.801
	Medium Human Development	
47	Saint Kitts and Nevis	0.798
48	Costa Rica	0.797
49	Croatia	0.795
50	Trinidad and Tobago	0.793
51	Dominica	0.793
52	Lithuania	0.789
53	Seychelles	0.786
54	Grenada	0.785
55	Mexico	0.784
56	Cuba	0.783
57	Belarus	0.781
58	Belize	0.777
59	Panama	0.776
60	Bulgaria	0.772
61	Malaysia	0.772
62	Russian Federation	0.771
63	Latvia	0.771
64	Romania	0.770
65	Venezuela	0.770
66	Fiji	0.769
67	Suriname	0.766
68	Colombia	0.764
69	Macedonia, TFYR	0.763
70	Georgia	0.762
71	Mauritius	0.761
72	Libyan Arab Jamahiriya	0.760
73	Kazakhstan	0.754
74	Brazil	0.747
75	Saudi Arabia	0.747
76	Thailand	0.745
77	Philippines	0.744
78	Ukraine	0.744
79	Saint Vincent and the Grenadines	0.738
80	Peru	0.737
81	Paraguay	0.736
82	Lebanon	0.735

Rank		Value (1998)
83	Jamaica	0.735
84	Sri Lanka	0.733
85	Turkey	0.732
86	Oman	0.730
87	Dominican Republic	0.729
88	Saint Lucia	0.728
89	Maldives	0.725
90	Azerbaijan	0.722
91	Ecuador	0.722
92	Jordan	0.721
93	Amenia	0.721
94	Albania	0.713
95	Samoa (Western)	0.711
96	Guyana	0.709
97	Iran, Islamic Rep. of	0.709
98	Kyrgyzstan	0.706
99	China	0.706
100	Turkmenistan	0.704
101	Tunisia	0.703
102	Moldova, Rep. of	0.700
103	South Africa	0.697
104	El Salvador	0.696
105	Cape Verde	0.688
106	Uzbekistan	0.686
107	Algeria	0.683
108	Viet Nam	0.671
109	Indonesia	0.670
110	Tajikistan	0.663
111	Syrian Arab Republic	0.660
112	Swaziland	0.655
113	Honduras	0.653
114	Bolivia	0.643
115	Namibia	0.632
116	Nicaragua	0.631
117	Mongolia	0.628
118	Vanuatu	0.623
119	Egypt	0.623
120	Guatemala	0.619
121	Solomon Islands	0.614
122	Botswana	0.593
123	Gabon	0.592
124	Morocco	0.589
125	Myanmar	0.585
126	Iraq	0.583
127	Lesotho	0.569

Rank		Value (1998)
128	India	0.563
129	Ghana	0.556
130	Zimbabwe	0.555
131	Equatorial Guinea	0.555
132	Sao Tome and Principe	0.547
133	Papua New Guinea	0.542
134	Cameroon	0.528
135	Pakistan	0.522
136	Cambodia	0.512
137	Comoros	0.510
138	Kenya	0.508
139	Congo	0.507
Low Human Development		
140	Lao People's Dem. Rep.	0.484
141	Madagascar	0.483
142	Bhutan	0.483
143	Sudan	0.477
144	Nepal	0.474
145	Togo	0.471
146	Bangladesh	0.461
147	Mauritania	0.451
148	Yemen	0.448
149	Djibouti	0.447
150	Haiti	0.440
151	Nigeria	0.439
152	Congo, Dem. rep. of the	0.430
153	Zambia	0.420
154	Cote d'Ivoire	0.420
155	Senegal	0.416
156	Tanzania, U. rep. of	0.415
157	Benin	0.411
158	Uganda	0.409
159	Eritrea	0.408
160	Angola	0.405
161	Gambia	0.396
162	Guinea	0.394
163	Malawi	0.385
164	Rwanda	0.382
165	Mali	0.380
166	Central African Republic	0.371
167	Chad	0.367
168	Mozambique	0.341
169	Guinea-Bissau	0.331
170	Burundi	0.321

Rank		Value (1998)
171	Ethiopia	0.309
172	Burkina Faso	0.303
173	Niger	0.293
174	Sierra Leone	0.252
All Developing Countries		**0.542**

* Source: Human Development Report 2000 UNDP (www.UNDP.org/hdr2000/English/hdr2000.html)

ANNEX 5:
DATA TEMPLATE FILE IN INFORMATION
File Information (F1household.sav)

List of variables on the working file

Name		Position
DATE	date of intervview Measurement Level: Ordinal Column Width: 8 Alignment: Right Print Format: F6 Write Format: F6	1
HHID	household identification Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F3 Write Format: F3	2
MFICLUST	MFI cluster code Measurement Level: Scale Column Width: 6 Alignment: Right Print Format: F1 Write Format: F1	3
GROUP	group name Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F2 Write Format: F2	4
LOCALITY	name of locality Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F2 Write Format: F2	5
MFICLIEN	Client of mfi Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F8.2 Write Format: F8.2 Value Label .00 no 1.00 yes	6

```
HHLDREPL  whether household iis from replacement list                    7
          Measurement Level: Ordinal
          Column Width: 8  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              0    no
              1    yes

ORIGHHLD  what the original household was                                8
          Measurement Level: Ordinal
          Column Width: 8  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              1    not found
              2    unwilling to answer
              3    client status was wrongly classified

RESPO     name of respondent                                             9
          Measurement Level: Nominal
          Column Width: 20  Alignment: Left
          Print Format: A20
          Write Format: A20

HHEAD     name of hhold head                                            12
          Measurement Level: Nominal
          Column Width: 20  Alignment: Left
          Print Format: A20
          Write Format: A20

HHADDRES  household address                                             15
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F5
          Write Format: F5

TOWN      address town                                                  16
          Measurement Level: Nominal
          Column Width: 20  Alignment: Left
          Print Format: A20
          Write Format: A20

_


INTERV    interviewer                                                   19
          Measurement Level: Ordinal
          Column Width: 8  Alignment: Right
          Print Format: F2
          Write Format: F2
```

```
          Value    Label

              1    Elliot
              2    Gitau
              3    Morris
              4    lavinia
              5    Betty
              6    Titus
              7    Alex
              8    Juliet

EVENT2DY  Was there a special event in past two days                    20
          Measurement Level: Ordinal
          Column Width: 8  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              0    no
              1    yes

MEALS2DY  Number of meals served in past 2 days                         21
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F2
          Write Format: F2

EVEMEAL2  number of meals served in 2 days preceding special event      22
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F2
          Write Format: F2

SPEVENWK  special event in the past week                                23
          Measurement Level: Ordinal
          Column Width: 8  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              0    no
              1    yes

_


LUXFOOD1  Number of days luxury food 1 served                           24
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F1
          Write Format: F1
```

LUXFOOD2 Number of days luxury food 2 served 25
Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

LUXFOOD3 Number of days luxury food 3 served 26
Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

INFERIOR Number of days inferior food served 27
Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

FOODMNTH Number of days in past month household members did not have 28
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F2
Write Format: F2

FOODYEAR Number of months in past year houshold members did not have 29
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F2
Write Format: F2

STAPLE1 frequency of purchasing staple 1 30
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value	Label
1	daily
2	twice a week
3	weekly
4	fortnightly
5	monthly
6	less frequently than a month
7	none

STAPLE2 frequency of purchasing staple 2 31
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value Label

1 daily
2 twice a week
3 weekly
4 fortnightly
5 monthly
6 less frequently than a month
7 none

STAPLE3 frequency of purchasing staple 3 32
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value Label

1 daily
2 twice a week
3 weekly
4 fortnightly
5 monthly
6 less frequently than a month
7 none

STOCKWKS number of weeks stock of local staple will last 33
Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F2
Write Format: F2

EXTRAMON amount of extra monthly income spent on food 34
Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F8.2
Write Format: F8.2

HOWNSHIP ownership status of house 35
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value Label

1 built on squatter land
2 given by relative or other to use
3 provided by government
4 rented
5 owned

NUMROMS number of rooms available 36
Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F2
Write Format: F2

ROOFTYPE type of roofing material used 37
Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value	Label
1	tarpaulin/plastic sheets/branchesand twigs
2	grass
3	stone/slate
4	iron sheets
5	brick tiles
6	concrete

WALLTYPE type of exterior walls 38
Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value	Label
1	tarpaulin/plastic sheets/branches and twigs
2	mud walls
3	iron sheets
4	timber
5	brick or stone with mud
6	brick or stone with cement plaster

—

HSECONDI structural condition of house 39
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value	Label
1	seriously dilapidated
2	need for major repairs
3	sound structure

ELECSUPP Household's electricity supply 40
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

```
          Value    Label

              1    no connection
              2    connection shared with others
              3    own metered connection

COOKFUEL  type of cooking fuel used                                    41
          Measurement Level: Nominal
          Column Width: 8  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              1    dung
              2    collected wood
              3    purchased wood or saw dust
              4    charcoal
              5    kerosene
              6    gas
              7    electricity

_


ACCWATER  quality of drinking water                                    42
          Measurement Level: Ordinal
          Column Width: 8  Alignment: Right
          Print Format: F2
          Write Format: F2

          Value    Label

              1    rain water
              2    dam
              3    pond/lake
              4    river/stream
              5    spring
              6    public well open
              7    public well sealed with pump
              8    well in residence,yard plot
              9    piped public water
             10    bore hole in residence
             11    private borehole in neighbours

LATRINE   quality of latrine                                           43
          Measurement Level: Nominal
          Column Width: 8  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              1    no facility/bush/field
              2    shared pit latrine
              3    own pit latrine
```

```
              4    shared ventilated  improved pit latrine
              5    own improved latrine
              6    shared flash toilet
              7    own flush toilet

AREACULT  size of cultivated land-local units                          44
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F5.2
          Write Format: F5.2

AREAUNCU  size of uncultivated land-local units                        45
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F5.3
          Write Format: F5.3

_


VALCULTI  value of cultivated landholdings                             46
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F8.2
          Write Format: F8.2

VALUNCUL  value of uncultivated landholdings                           47
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F8.2
          Write Format: F8.2

MFIWEALT  relative wealth assessment of MFI clients                    48
          Measurement Level: Ordinal
          Column Width: 8  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              1    poor
              2    not poor
              3    rich
              4    dont know mfi

_
```

File Information (F2adult.sav)

List of variables on the working file

Name		Position
HHID	household identification	1

Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F3
Write Format: F3

MFICLUST MFI area code 2
Measurement Level: Nominal
Column Width: 6 Alignment: Right
Print Format: F1
Write Format: F1

GROUP MFI group name 3
Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F2
Write Format: F2

LOCALITY name of locality 4
Measurement Level: Nominal
Column Width: 6 Alignment: Right
Print Format: F2
Write Format: F2

MFICLIEN MFI client status 5
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value	Label
1	client of mfi
2	nonclient of mfi

MEMBERID ID code of hosuehold member 6
Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F2
Write Format: F2

NAME first name of household member 7
Measurement Level: Nominal
Column Width: 12 Alignment: Left
Print Format: A20
Write Format: A20

HHSTATUS status of the household head 10
Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value	Label
1	single
2	married with spouse permanently present in hhold
3	married with spouse migrant
4	widower
5	divorced/separated
6	living with other wife, contribution limited

RELATION relationship to head of household 11
Measurement Level: Nominal
Column Width: 6 Alignment: Right
Print Format: F1
Write Format: F1

Value	Label
1	head of hhold
2	spouse
3	son or daughter
4	father or mother
5	grand child
6	grand parents
7	other relative
8	other nonrelative

SEX sex of the household members 12
Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value	Label
1	male
2	female

—

AGE age of adult members 13
Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F3
Write Format: F3

MAXEDUCA max. level of schooling 14
Measurement Level: Ordinal
Column Width: 8 Alignment: Right

Print Format: F1
Write Format: F1

Value Label

1 less than primary
2 some primary
3 completed primary
4 attended polytechnic
5 attended secondary
6 completed secondary
7 attended college or university

CANWRITE Household member can write 15
Measurement Level: Ordinal
Column Width: 7 Alignment: Right
Print Format: F1
Write Format: F1

Value Label

0 no
1 yes

OCCUPAT Main occupation of household member 16
Measurement Level: Ordinal
Column Width: 7 Alignment: Right
Print Format: F1
Write Format: F1

Value Label

1 self employed in Agriculture
2 Self employed in nonfarm enterprise
3 pupil/student
4 casual
5 salaried worker
6 domestic work
7 unemployed, looking for a job
8 unwilling to work/retired
9 not able to work/handicapped

_

MFICURRE if current member of mfi 17
Measurement Level: Ordinal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1

Value Label

0 no
1 yes

AMTLOAN	Amount of loan borrowed from MFI Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F5.2 Write Format: F5.2	18
COTHEXPE	Expenditures on clothing and footwear Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F6.2 Write Format: F6.2	19

File Information (F3child.sav)

List of variables on the working file

Name		Position
HHID	household identification Measurement Level: Nominal Column Width: 8 Alignment: Right Print Format: F3 Write Format: F3	1
MFICLUST	MFI cluster code Measurement Level: Nominal Column Width: 6 Alignment: Right Print Format: F1 Write Format: F1	2
GROUP	MFI group name Measurement Level: Nominal Column Width: 8 Alignment: Right Print Format: F2 Write Format: F2	3
LOCALITY	name of locality Measurement Level: Nominal Column Width: 6 Alignment: Right Print Format: F2 Write Format: F2	4
MFICLIEN	Client of mfi Measurement Level: Nominal Column Width: 8 Alignment: Right Print Format: F8.2 Write Format: F8.2 Value Label .00 no 1.00 yes	5
IDCHILD	identification code for child in household Measurement Level: Nominal Column Width: 8 Alignment: Right Print Format: F2 Write Format: F2	6
NAME	name of child in household Measurement Level: Nominal Column Width: 10 Alignment: Left Print Format: A10 Write Format: A10	7

AGE	age of child Measurement Level: Scale Column Width: 3 Alignment: Right Print Format: F2 Write Format: F2	9
COTHEXPE	Expenditures on clothing and footwear Measurement Level: Scale Column Width: 7 Alignment: Right Print Format: F6.2 Write Format: F6.2	10

File Information (F4assets.sav)

List of variables on the working file

Name		Position
HHID	household identification code Measurement Level: Nominal Column Width: 8 Alignment: Right Print Format: F3 Write Format: F3	1
MFICLUST	MFI cluster code Measurement Level: Nominal Column Width: 6 Alignment: Right Print Format: F1 Write Format: F1	2
GROUP	MFI group name Measurement Level: Nominal Column Width: 8 Alignment: Right Print Format: F2 Write Format: F2	3
LOCALITY	name of locality Measurement Level: Nominal Column Width: 6 Alignment: Right Print Format: F2 Write Format: F2	4
MFICLIEN	Client of mfi Measurement Level: Nominal Column Width: 8 Alignment: Right Print Format: F8.2 Write Format: F8.2 Value Label .00 no 1.00 yes	5
ASSET	asset type Measurement Level: Nominal Column Width: 8 Alignment: Right Print Format: F2 Write Format: F2 Value Label 1 cattle and buffalo 2 sheep, goats and pigs 3 adult poultry and rabbits 4 horses and donkeys 5 cars	6

```
             6    motorcycles
             7    bicycles
             8    Other vehicles
             9    carts
            10    TVs
            11    VCRs
            12    refrigerators
            13    electric/gas cookers
            14    washing machines
            15    radios
            16    fans

NUMOWNED  Number owned by household                                    7
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F4
          Write Format: F4

RESALE    Resale value of asset at current market price               8
          Measurement Level: Scale
          Column Width: 8  Alignment: Right
          Print Format: F9.2
          Write Format: F9.2

_
```

BIBLIOGRAPHY

Aho, G., S. Larivière, and F. Martin. 1998. *Poverty analysis manual: Applications in Benin*. New York: United Nations Development Programme and Université Nationale du Bénin, Université Laval.

Basilevsky, A. 1994. *Statistical factor analysis and related methods.* New York: John Wiley and Sons, Inc.

Bennett, S., A. Radalowicz, V. Vella, and A. Tomkins. 1994. A computer simulation of household sampling schemes for health surveys in developing countries. *International Journal of Epidemiology* 23 (6): 1282–1291.

Carletto, C. 1999. *Constructing samples for characterizing household food security and for monitoring and evaluating food security interventions: Theoretical concerns and practical guidelines.* Technical Guide No. 8. Washington, D.C.: International Food Policy Research Institute.

Child, D. 1990. *The essentials of factor analysis*. Second edition. London: Cassell Educational Limited.

Chung, K., L. Haddad, J. Ramakrishna, and F. Riely. 1997. *Identifying the food insecure: The application of mixed-method approaches in India*. Washington, D.C.: International Food Policy Research Institute.

Gibbons, D. S., A. Simanowitz, and B. Nkuna. 1999. *Cost-effective targeting—Two tools to identify the poor: Operational manual*. CASHPOR-SEF.

Hatch, J., and L. Frederick. 1998. *Poverty assessment by micro-finance institutions: A review of current practice*. Bethesda, Md., U.S.A.: Micro-enterprise Best Practices.

Lipton, M., and M. Ravallion. 1995. Poverty and policy. In *Handbook of development economics*, Vol. IIIB, eds. J. Behrman and T. Srinivasan. Amsterdam: Elsevier.

Minten, B., and M. Zeller, eds. 2000. *Beyond market liberalization: Income generation, welfare and environmental sustainability in Madagascar*. Aldershot, U.K.: Ashgate Publishing Company.

Navajas, S., M. Schreiner, R. Meyer, C. Gonzalez-Vega, and J. Rodriguez-Meza. 2000. *Micro-credit and the poorest of the poor: Theory and evidence from Bolivia.* World Development Report. Washington, D.C.: World Bank, forthcoming.

Skoufias, E., B. Davis and H. Soto. 2000. Practical alternatives to consumption based welfare measures. photocopy.

Streeten, P. 1994. Poverty concept and measurement. In *Poverty monitoring: An international concern*, eds. R. van der Hoeven, and R. Anker, eds. New York: St Martin's Press.

UNDP (United Nations Development Programme). 1990, 1997, 1999. *Human development report*. New York: Oxford University Press.