



Assessing the Impact of Microenterprise Services (AIMS)

Management Systems International
600 Water Street, S.W.
Washington, D.C. 20024-2488
Tel: (202) 484-7170 • Fax: (202) 488-0754
E-mail: aims@msi-inc.com

RESEARCH STRATEGY FOR THE AIMS CORE IMPACT ASSESSMENTS

March 2002

Submitted to:

Monique Cohen, Ph.D.
Office of Microenterprise Development
Global Bureau, USAID

Submitted by:

Elizabeth Dunn, Ph.D.
Department of Agricultural Economics
Social Sciences Unit
University of Missouri-Columbia

This work was funded by the Microenterprise Impact Project (PCE-0406-C-00-5036-00) of USAID's Office of Microenterprise Development. The Project is conducted through a contract with Management Systems International, in cooperation with Harvard University, the University of Missouri, and the Small Enterprise Education and Promotion Network.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
EXECUTIVE SUMMARY	v
I. INTRODUCTION	1
A. Background on the AIMS Core Impact Assessments.....	1
B. Design Challenges	2
C. Conceptual Model and Hypotheses	3
II. SELECTION OF THE RESEARCH DESIGN	7
A. The Impact Problem.....	7
B. Selection Bias in Impact Assessment	7
C. Experimental and Non-Experimental Designs.....	8
D. Use of Panel Data in Non-Experimental Designs.....	9
E. Key Design Features of the CIA.....	12
III. COLLECTION AND ANALYSIS OF PANEL SURVEY DATA.....	13
A. The CIA Panel Data	13
B. Measurement of Core Impact Variables	14
C. Sample Design and Data Collection	15
D. Data Analysis	18
IV. COLLECTION AND ANALYSIS OF CASE STUDY DATA	24
A. Objectives of the Case Study Research	24
B. Study Questions and Propositions	25
C. Implementation Issues	26
D. Analyzing the Case Study Database	28
V. SUMMARY AND CONCLUSION	29
A. Advantages of the Research Strategy	29
B. Disadvantages of the Research Strategy and Implications for Future Research	31
C. Conclusion	34
REFERENCE LIST	35
APPENDIX: Calculating Gain Scores.....	44

LIST OF TABLES

Table 1: Household-Level Impact Variables	40
Table 2: Enterprise-Level Impact Variables	41
Table 3: Individual-Level Impact Variables	42
Table 4: Study Questions and Research Propositions for Case Studies	43

LIST OF FIGURES

Figure 1: Conceptual Model of the Household Economic Portfolio	39
Figure 2: Non-Experimental Panel Design	10
Figure 3: Example of Changes in Household Income in Non-Experimental Panel Data	11
Figure 4: Example of Gain Scores for Nominal Data	45

ACKNOWLEDGEMENTS

The core impact assessments were the result of several years of close collaboration among a team of researchers on the AIMS Project. In particular, the research strategy described in this paper was developed in collaboration with the other three primary investigators for the core impact assessments. Carolyn Barnes of Management Systems International provided extensive input throughout the entire design process. Donald Snodgrass and Martha Chen of Harvard University were also actively involved in the research design. Special recognition is due to Jennefer Sebstad and Monique Cohen, who provided critical review and substantive input at each stage of the research. In addition, team members Elaine Edgcomb and Russ Webster provided creative input during the design and implementation of the core impact assessments.

This paper represents the consolidation of several published and unpublished documents describing specific design features of the core impact assessments. Each of these design components was externally reviewed and revised based on the comments of the external reviewers. In the initial stages, a number of people contributed to discussions on the research plan, including Bruce Bolnick, Kirk Dearden, Gary Gaile, Jonathan Morduch, and Joan Parker. Andrew McKay and Ping Yu reviewed the first written version of the research plan. Anthony Turner and Robert Magnani reviewed the sampling design for the longitudinal survey. David Hulme, Nancy Horn, and Ron Chua reviewed the reports on the baseline findings. Charles Reichardt reviewed the plan for the analysis of the survey data as well as the drafts of the final impact results. The final impact reports were also reviewed by Elaine Edgcomb, Mark Schreiner, and Peter Little. The contributions of all of these external reviewers are gratefully acknowledged.

Finally, the author would like to acknowledge the helpful comments provided by Carolyn Barnes and Donald Snodgrass in their review of this paper and the editorial assistance of J. Gordon Arbuckle, Jr. Any remaining errors or omissions, however, remain the responsibility of the author.

EXECUTIVE SUMMARY

The Assessing the Impacts of Microenterprise Services (AIMS) Project was sponsored by USAID between 1995 and 2002, as part of the agency's Microenterprise Innovation Project. The core impact assessments (CIA) were a key component of the AIMS Project. The CIA were methodologically rigorous, longitudinal impact studies of three microenterprise programs: SEWA Bank in India, Mibanco in Peru, and Zambuko Trust in Zimbabwe. All three of these programs provide microenterprise credit, with the SEWA program also offering savings services and non-enterprise credit and the Zambuko program providing business training. The three studies grew out of a common research design and tested the same set of core impact hypotheses using a similar research approach. The purpose of this paper is to describe the common methodological approach used in the three CIA studies.

Design Challenges

For an impact assessment to produce useful and credible results, there are a number of significant conceptual challenges that it must address. Two of these challenges, namely attribution and selection bias, are generic in the sense that they threaten the validity of any impact study in the social sciences. The CIA studies were designed to address both of these challenges. They were also designed to address the issue of fungibility, which relates to credit and money.

Fungibility is a basic characteristic of money. It means that monetary units are interchangeable and can be used for a wide variety of purposes. Moreover, it is difficult to trace how a household allocates money, including the money provided in the form of loan funds. Without knowing how much of a loan reaches the intended enterprise, it is difficult to link the receipt of the loan to the changes observed in the enterprise.

Establishing a strong, plausible case for attribution is another conceptual challenge. Statistical methods can establish statistical correlation, but they cannot prove the existence of a cause-and-effect relationship. In addition, controlled experiments in which all factors except the treatment are held constant are very difficult to conduct in the social sciences. The result is that it can never be proven beyond doubt that a treatment leads to an impact. Instead, the best that can be done in social science impact evaluation is to establish a strong, plausible case for attribution.

A third design challenge is selection bias. Selection bias stems from the fact that people self-select whether or not they will apply for microenterprise services; program managers and credit agents also select service areas and individual clients on the basis of their likelihood of succeeding. Selection bias can exaggerate the findings of an impact assessment because the observed differences in the impact variables may be due either to the impact of participation in the program or to inherent differences in the people who entered the program.

The Conceptual Model

Threats to internal validity stemming from fungibility and attribution can be reduced through the use of a conceptual framework that explicitly addresses these challenges. The problem of fungibility can be addressed by widening the unit of analysis for the impact assessment from a

single enterprise to the entire economic portfolio within which the fungible capital might be used. This eliminates the need to assume that all of the loan funds were spent on the intended enterprise. The problem of attribution can be addressed with an internally consistent conceptual framework that models the ways that households, and individuals within households, use credit to protect, manage, and increase their resources and activities, including their microenterprises.

The household economic portfolio (HHEP) model is suitable for addressing these challenges because it is based on a conceptualization of the microenterprise as part of a larger portfolio of household economic activities. Decisions about microenterprises are made in the context of options and tradeoffs within the overall household economy. Microenterprise credit is modeled as a fungible addition to household resources and decision makers within the household are assumed to allocate the loan funds to the activities they consider most important.

The HHEP model provides a framework for developing hypotheses about the cause-and-effect relationships between microenterprise services and impacts. The CIA research in all three countries was designed to test a set of common, or “core” impact hypotheses at three levels. These hypotheses posit that microenterprise services lead to impacts, or changes, at the household, the microenterprise, and the individual levels.

SELECTION OF THE RESEARCH DESIGN

The basic challenge in impact assessment is to determine the effect of an intervention on an outcome variable. In other words, impact evaluation seeks to measure the difference in outcome between an individual who received the treatment and what the outcome *would have been* for the same individual, if he or she *had not received* the treatment. Since the latter is an unobservable event, the only practical alternative is to compare the outcome for individuals who receive treatment with the outcome for individuals who do not receive treatment. This is a fundamental challenge in impact evaluation and the source of the associated selection bias problem.

The selection bias problem implies that individuals who receive treatment and those who do not may be inherently different, and that these differences may lead to incorrect measurement of the treatment effect. In this case, the program participants all chose to receive the “treatment” by becoming program clients. It is possible that clients differ, on average, from those who choose not to participate. If differences between participants and non-participants relate to the ability to realize benefits from program services, that could lead to differences in the outcome variables (e.g., income and revenue) that should not be attributed to the program.

In a non-experimental design, also known as a quasi-experimental design, the outcome variable is measured for the treatment group and for a constructed control group of respondents who do not receive the treatment but who are similar to the treatment group in critical ways that affect outcomes. The most commonly used method for constructing a control group is to select respondents who share critical characteristics with the treatment group, then to control statistically for differences in other variables that are expected to affect outcomes. Panel data, which follows the same respondents over time, can help to address the selection bias problem by accounting for the fixed effects of selection bias and for the exogenous effects on outcomes that are unrelated to program participation.

The purpose of the AIMS core impact assessments was to generate strong, plausible inferences about the impacts of microenterprise services on clients, their enterprises, and their households. In order to do so, the studies relied on a mixed-method approach. This approach combined quantitative and qualitative methods to reach a new level of understanding about the clients of microenterprise programs, the positive and negative impacts of those programs, and the magnitude of those impacts.

COLLECTION AND ANALYSIS OF PANEL SURVEY DATA

The quantitative component of the CIA was based on a non-experimental research design utilizing panel data. The use of panel data is an important strength of the CIA studies, allowing greater simplicity and transparency in the choice of analytical methods. More specifically, the CIA data correspond to a prospective panel design in which the survey occurs at two points in time. The same respondents were interviewed both times, and the same operational variables were measured each time.

Sample Selection and Data Collection

In each of the sites, a two-stage sampling approach was followed. First, geographical regions were selected, then client households were randomly selected from client lists provided by the microenterprise support programs. Comparable non-client households were randomly selected from within the same neighborhoods as the clients. The non-clients were similar to the clients in terms of gender, sector, and location characteristics and their eligibility for program participation. In Peru and Zimbabwe, the CIA survey data were collected in 1997 and 1999. The India survey data were collected in 1998 and 2000. In all three countries, the two survey rounds occurred at the same time of the year.

The client sample (the treatment group) included households who were classified as clients at the time of the baseline survey. Specifically, the *client sample* included 1) all households who were clients of the program in both the baseline and second-round periods and 2) all households who were clients during the baseline, but who were no longer clients at the time of the second-round survey. The *non-client sample* (the control group) consisted of all households who had never received services from the program being studied.

In each of the three studies, there were subgroups of the client sample that were defined in terms of their length, level, or type of program participation. For the study in India, there were two client samples: 1) the *borrowers*, who had SEWA loans and SEWA savings at the time of the baseline survey and 2) the *savers*, who had SEWA savings, but no loan, at the time of the baseline. In Peru, special treatment was given to respondents who received their first microenterprise loan between the first and second rounds of the survey. The separate analysis of this *new entrant group* provided information about the size of the treatment effect relative to the size of selection bias. In Zimbabwe, the clients were grouped for separate analysis on the basis of whether or not they continued to receive loans after the baseline survey. In addition, some distinctions were made between clients who had received only one loan at the time of the baseline and clients who had already received more than one loan at that time.

Data Analysis

The survey data were analyzed using several complementary approaches. In order to provide information on changes in the outcome variables between the two rounds of the survey, paired t-tests and gain score analysis were used. The results of the paired t-tests and gain score analysis were treated as part of the descriptive analysis, providing background information about the size and direction of changes in the outcome variables for each of the comparison groups. Analysis of covariance (ANCOVA) was the central approach used to analyze the panel data and test the hypotheses about the impacts of microenterprise services.

For each impact variable, there were key moderating variables believed to affect the relationship between program participation and the level of change in the impact variable. Some of the moderating variables used in the analysis were gender of the entrepreneur, enterprise sector, and household composition. Moderating variables were included in the analysis for several related reasons: 1) if the treatment and control groups differ significantly in the distribution of these moderating variables, then at least some of the measured differences in the impact variables may be due to differences in the moderating variables rather than to program participation; 2) by including the moderating variables, it may be possible to establish a statistical relationship between the outcome variable and the moderating variables; and 3) moderating variables may help to statistically explain variations in outcomes, thus providing a more powerful and precise analysis of impact.

The ANCOVA procedure has the advantage of statistically controlling for multiple differences between the treatment and control groups, differences that may affect the relationship between program participation and changes in the outcome variables. In effect, the ANCOVA procedure statistically “matches” individuals in the treatment and control groups who have the same baseline measures on the outcome variable and similar values of the moderating variables. It then compares these matched observations to determine if there are any consistent differences between the treatment and control groups in terms of their outcome values in the second round. Since ANCOVA adjusts the estimate of the treatment effect to account for differences in the baseline measures and moderating variables, it reduces the influence of selection bias on the estimates of impact.

COLLECTION AND ANALYSIS OF CASE STUDY DATA

In order to develop an in-depth understanding of program impact, the core impact assessments included case study research. The case studies were designed to augment the survey data by examining how and why changes occur as the result of program participation. The case study research supplemented the survey by following a case-within-survey design with pre-program and post-program impact measures. Using the household economic portfolio model as a conceptual framework, the qualitative research focused on enterprise, household, and individual-level variables to address the following research question: “How do microfinance programs contribute to the observed changes within the household and its enterprises?” Through reconstruction of the chain of events leading from program participation to the measured impacts, the case study data complement and strengthen the survey results and help to strengthen the case for attribution.

The case study research included nine to twelve case study households in each country, selected on the basis of level of program participation and additional relevant variables, such as household income or asset level (India and Peru), gender (Peru), trade group (India), and program (Zimbabwe). Multiple cases were selected in each subgroup to provide literal replication. Theoretical replication was expressed in terms of differences in impacts across the subgroups (e.g., impacts for new participants differing from impacts for long-term participants). All of the households selected into the case studies were program clients.

There were two rounds of case study interviews, one year apart. The first round of interviews focused on changes that had occurred since the client joined the program, with particular emphasis on the changes that had occurred in the period immediately before and after joining the program. The second round of case study interviews, conducted with the same households that participated in the first round, focused on changes that occurred between the first and second rounds of the survey. The data were analyzed using a variant of pattern matching in which the patterns in the empirical evidence were compared to patterns predicted in the study propositions. If the patterns matched, that was taken as evidence in favor of the study propositions.

SUMMARY AND CONCLUSION

The three core impact assessments produced a relatively rigorous set of impact evaluations that point the way toward improved methodology in the field. These studies established some strong, plausible evidence for the impacts of microenterprise services and showed that impacts are not always where one might expect to find them. There were four major design features that contributed to the strength of the research strategy:

- By widening the unit of analysis to include impacts on the entire household economy, the conceptual framework addressed the problem of fungibility and provided a logical framework for attributing the observed impacts to the program services received.
- The mixed-method approach combined quantitative and qualitative data to yield a much more informed view about how and why impacts occur and to strengthen the case for attribution.
- The quasi-experimental design controlled for the influence of external, non-program factors that affect the outcomes for both clients and non-clients, thus establishing the underlying trends related to the counterfactual. The fact that the non-clients in the control group were eligible for program participation helped to reduce selection bias and improve the case for attribution.
- The use of panel data, along with a statistical approach that incorporated information on the starting values of the impact variables, helped to reduce some of the influence of selection bias on the impact results.

The research approach also has several important limitations:

- Selection bias may not have been entirely eliminated. The ANCOVA approach does not control for differences on unobserved variables that may affect outcomes, possibly leading to overestimation of positive impacts and underestimation of negative impacts.
- The baseline measures might already reflect impacts because they are not true pre-treatment measures of the outcome variables, possibly leading to underestimation of both positive and negative impacts. While selection bias and the lack of a pre-treatment baseline may have opposite effects on positive impacts, they both may lead to underestimation of negative impacts.
- The research approach relied on relatively unsophisticated measures of program participation and there were measurement weaknesses in some of the impact indicators.

The methodological limitations of the AIMS CIA research strategy should be kept in mind when interpreting the research findings and planning future studies. However, many of these limitations are the results of practical considerations and tradeoffs among options. For example, selection bias could be eliminated entirely through the use of an experimental design, which is the gold standard in social science impact evaluation. However, the random selection of qualified applicants to receive or not receive program services raises ethical and public relations issues and is generally rejected by managers of microenterprise programs. Similarly, the use of alternative econometric procedures to remove some of the effects of selection bias would introduce additional assumptions and limitations into the research.

The use of a true pre-treatment baseline might be another improvement. With a pre-treatment baseline, only incoming clients would be included in the treatment group. However, this type of study would need to last for five or more years in order to generate information on longer term impacts. Future impact studies could also benefit from more sophisticated measures of program participation, but these are difficult to construct with the types of data current available in clients' credit history files. Finally, there is a need for impact assessments that look beyond the household to market-level, regional, and macroeconomic impacts, levels of analysis that were beyond the scope of the AIMS studies. By combining this information with the detailed household- and enterprise-level results of studies like the CIA, it would be possible to develop a comprehensive understanding of the costs and benefits of microenterprise programs and to improve the efficacy of these programs in achieving economic development objectives.

I. INTRODUCTION

A. Background on the AIMS Core Impact Assessments

Between 1995 and 2002, USAID sponsored the Assessing the Impacts of Microenterprise Services (AIMS) Project as part of the Microenterprise Innovation Project. An overall goal of the AIMS Project was to gain a better understanding of the processes by which microenterprise services strengthen businesses and improve the welfare of microentrepreneurs and their households. Another goal was to improve the ability of USAID and its partners to assess the impacts of their microenterprise programs and to generate client-level information that could be useful for program management. The project was designed to develop a practical, yet conceptually grounded, approach to measuring the impacts of microenterprise services on enterprise growth and on the well being of clients and their households.

There were three major research components in the AIMS Project. Under the action research component, a series of topics related to impact assessment and client-level behavior were addressed in both desk studies and field-focused research. The second component centered on the development of a set of low-cost, credible tools for practitioners to use in measuring client-level impacts and client satisfaction.¹ This paper describes the research strategy for a third research component, namely the AIMS core impact assessments.

The AIMS core impact assessments, or CIA, are methodologically rigorous, longitudinal impact studies of three microenterprise support programs: SEWA Bank in India, Mibanco in Peru, and Zambuko Trust in Zimbabwe. These three programs were selected on the basis of several criteria. First, they are geographically dispersed, representing programs in Asia, Latin America, and Africa. In addition, all three programs had been offering microenterprise credit for several years under a stable methodology, and their financial reports indicated that the programs were at or near operational self-sufficiency. The clients of the three organizations also fell along a poverty continuum, with SEWA's clients being the poorest, Zambuko's clients in the middle of the range, and Mibanco's clients mostly above the poverty line.

The three impact assessments in the CIA were designed to test the same set of core impact hypotheses and followed a similar research approach. The baseline studies and final impact results have been published under the AIMS Project.² These papers focus on the research findings and include only partial descriptions of the research approach. Until now, the full description of the research approach for the CIA could only be found in a series of externally reviewed, but unpublished, documents. This paper seeks to fill that gap and pull the information together in one document. The objectives of this paper are as follows: 1) to document and justify the research procedures and methods used in the CIA; 2) to synthesize key elements of unpublished planning papers in order to make information on the CIA research strategy publicly

¹ The conceptual papers, field studies, and tools manual for practitioners can be downloaded from the project web site (www.mip.org). In addition to the three research components, the AIMS Project also included mission technical assistance and information dissemination components.

² The reference list at the end of this document provides the full citations for the baseline reports (Chen and Snodgrass 1999; Barnes and Keogh 1999; Dunn 1999) and for the final impact studies (Chen and Snodgrass 2001; Barnes 2001; Dunn and Arbuckle 2001), which are also available on the project web site (www.mip.org).

accessible; and 3) to provide an example of the process for implementing a methodologically rigorous impact assessment.

B. Design Challenges

There are a number of conceptual challenges associated with conducting impact assessments in the social sciences. These challenges affect not only the design of impact assessments but, ultimately, the usefulness and credibility of the results. While the design challenges may appear formidable, they should not be considered so intractable as to preclude any attempt to measure the impacts of microenterprise services. Instead, there are ways to at least partially address these inherent difficulties. This section introduces three specific design challenges: fungibility, attribution, and selection bias. Much of the rest of the paper describes how, and to what extent, these challenges are addressed by the research design.

1. Fungibility

Fungibility is a basic characteristic of money: it means that monetary units are interchangeable and can be used for a wide variety of purposes. The money that a client receives as a program loan may be used within the intended microenterprise, but some or all of the money might also be used in a different microenterprise or used to purchase property, pay school fees, or buy groceries. Not only can money be used in a variety of ways, it is usually very difficult to accurately trace how a household allocates loan funds.³

In the past, the fungibility of money was considered a major impediment to conducting impact assessments of microenterprise services. According to this view, the fact that a client may not spend all of the loan proceeds on the intended microenterprise meant that an unknown amount of loan funds were reaching the microenterprise. In addition, the household might mobilize other (non-program) funds to support the enterprise. This made it difficult to link any changes in the enterprise to the financial services received, thus undermining the premise of impact assessment.

The problem of fungibility can be addressed by widening the unit of analysis for the impact assessment. Instead of focusing only on the microenterprise, the unit of analysis should be expanded to include the entire economic portfolio within which the fungible capital might be used. In this way, there is no longer a need to assume that all of the loan funds are spent on the intended enterprise or, alternatively, to attempt to track how loan funds are actually spent.

2. Attribution

Another important conceptual challenge in evaluating the impact of microenterprise services is the problem of establishing a strong, plausible case for attribution. The attribution problem is a general one, affecting impact evaluation in all the social sciences. Basically, the problem arises from two causes. First, the statistical methods used to measure impacts can establish statistical

³ The issue of additionality is separate from, but related to, fungibility. Additionality occurs when the loan leads to a net increase in microenterprise investment that would not have occurred in the absence of the loan. There is no additionality if the same microenterprise investment would have been made in the absence of the loan, through the reallocation of (fungible) money away from some other use within the household economic portfolio.

correlation, but they cannot be used to prove the existence of a cause-and-effect relationship. Second, controlled experiments in which all factors except the treatment (intervention) are held constant are very difficult to conduct in the social sciences. The result is that it can never be proven incontrovertibly that the treatment led to the impact. Instead, the best that can be done is to establish a strong case in favor of attribution.

One of the ways to build a plausible case for attribution is to base the research on an internally consistent conceptual model that links the intervention to the impact in a plausible cause-and-effect relationship. The household economic portfolio model, described below, provides just such a conceptual framework. It can be used to model the ways that households, and the individuals within households, use microenterprise services to protect, manage, and increase their resources and activities, including their microenterprises.

Another way to strengthen the case for attribution is to use qualitative research to identify and document the chain of events, ordered in time, that lead from the program services to the impact. Qualitative research can also be used to eliminate the possibility of alternative explanations for the impacts that are observed. Any one of these approaches--qualitative evidence, a logically consistent conceptual model, statistical results indicating association between variables--could be used alone to build a plausible case for attribution. By combining them, however, the mixed-method approach builds a much stronger case for attribution.

3. Selection Bias

A third important challenge to the design of an impact assessment is selection bias. Selection bias can occur for two reasons. First, it can occur because people self-select whether or not they will apply for microenterprise services. Selection bias can also occur because program managers select service areas on the basis of the probability of program success and credit agents select individual clients on the perceived creditworthiness of these potential clients.

The problem with selection bias is that it can exaggerate the measured impacts. When samples of clients and non-clients are drawn, the observed differences in outcomes might be due either to the impact of participation in credit program (i.e., the “true” impact) or the differences might be due to the fact that the people who applied for credit were already different in some way(s) that would have led them to have greater success anyway. The selection bias problem is discussed in more detail in section II of this paper. That section focuses on the role of the research design in helping to reduce selection bias.

C. Conceptual Model and Hypotheses

The conceptual basis for the CIA is described in *Conceptual Framework for Assessing the Impacts of Microenterprise Services* (AIMS Team 2002), a paper that synthesizes earlier AIMS publications. In particular, there were two earlier AIMS papers, that contributed significantly to the conceptual framework (Sebstad et al. 1995; Chen and Dunn 1996). This section briefly summarizes the conceptual framework and lists the specific impact hypotheses that were tested in the three CIA studies. For additional details, however, the reader is referred to these other papers that focus specifically on the conceptual framework.

1. The Household Economic Portfolio Model

Sebstad et al. (1995) describe a framework for assessing the impacts of microenterprise interventions. This preliminary framework outlines a causal path through which microenterprise interventions lead to impacts. An important contribution of this framework is that it is based on the conceptualization of the microenterprise as part of a larger portfolio of household economic activities, with decisions about microenterprises being made in the context of options and tradeoffs within the overall household economy.

In the conceptual model of the household economic portfolio, the household is defined in terms of three components: 1) the human, physical, and financial resources of the household; 2) the production, consumption, and investment activities of the household; and 3) the circular flows between resources and activities (Chen and Dunn 1996). These circular flows include both the decisions that allocate resources to activities and the return flow of income and other resources generated by the selected activities. This return flow of income serves to augment the set of household resources.

The microenterprise is only one of several activities embedded in the household economy and supported by the same set of household resources. The microenterprise supported by the program may be one of several income-generating activities that draw on the household's limited resources. Microenterprise credit is considered a fungible addition to household resources that can be allocated to the activity (or activities) considered most important by the individuals within the household who control the credit allocation decision. A graphical representation of the household economic portfolio model is provided in figure 1 (at the end of the document).

2. Hypotheses

The CIA research in all three countries was guided by a set of common, or "core" impact hypotheses at the household, the microenterprise, and the individual (entrepreneur) levels. These core hypotheses were based on the conceptual model described above and were refined and finalized following 1) a review of prior microenterprise impact evaluations (Sebstad and Chen 1996); 2) pilot field investigations at the three core impact assessment sites in India, Peru, and Zimbabwe; and 3) a series of discussions and working meetings among the members of the AIMS team and with outside experts.

In addition to the core hypotheses, which were tested in all three CIA studies, a limited number of supplementary hypotheses were identified for each country. These supplementary hypotheses focused on impacts that were considered relevant to the specific context within which the assessment took place, such as program mission and emphasis, unique program services, cultural setting, or macroeconomic conditions.

The impact hypotheses posit that participation in microenterprise services leads to impacts at three levels: the household, the enterprise, and the individual client. Within each of these levels, a number of possible impacts are hypothesized. The core impact hypotheses are listed below.

Impacts at the household (H) level:

- H-1. an increase in the level of household income;
- H-2. greater diversification in the sources of household income;
- H-3. an increase in household assets, including
 - (H-3a) improvements in housing,
 - (H-3b) increases in major household appliances and transport vehicles, and
 - (H-3c) increases in microenterprise fixed assets;⁴
- H-4. an increase in expenditures on children's education;
- H-5. an increase in expenditures on food, especially among the very poor; and
- H-6. an increase in the household's effectiveness in coping with shocks.

Impacts at the enterprise (E) level:

- E-1. an increase in microenterprise revenue;
- E-2. an increase in enterprise fixed assets, especially among repeat borrowers;
- E-3. an increase in the paid and unpaid employment generated by the enterprise; and
- E-4. improvements in the transactional relationships of the enterprise.

Impacts at the individual (I) level:

- I-1. an increase in the client's control over resources and income within the household economic portfolio;
- I-2. increased self-esteem and respect from others;
- I-3. an increased incidence of personal savings; and
- I-4. a better position from which to deal with the future through more proactive behavior and increased confidence.

By sharing a set of core hypotheses, the studies in all three countries focused on the same conceptual variables. However, the techniques used to measure each variable sometimes differed across the three countries. For example, hypotheses H-4 and H-5 refer to expenditures on children's education and expenditures on food, respectively. In some countries, direct measurements of expenditures were used. In other countries, proxy measures were used, such as school enrollment as a proxy for education expenditures or frequency of eating certain foods as a proxy for expenditures on food. The measurement of the impact variables is discussed in more detail in section III.

It should be noted that the specific microenterprise services being investigated differed by country. In all three countries, the collaborating programs offered microenterprise credit. In addition, some business training was provided in Zimbabwe, while the program in India offered more comprehensive services, including savings services, insurance, and credit for purposes other than microenterprises. These programmatic differences were expected to have an influence on the pattern of impacts observed in each of the three studies.

⁴ This household-level hypothesis refers to the total (aggregate) value of fixed assets for all enterprises in the household. The enterprise-level hypothesis (E-2) refers to the fixed assets of a single enterprise.

3. Role of the Conceptual Model in Addressing Design Challenges

The household economic portfolio model is useful for addressing the issues of attribution and fungibility. Credit is fungible within the household economic portfolio in the sense that it is interchangeable with other monetary units and difficult to trace. The conceptual model recognizes that loan funds, like any of the household's resources, can be allocated to any activity in the household economic portfolio. The microenterprise is embedded in the household economy and represents only one of the household's production, consumption, and investment activities. By treating the microenterprise as part of the larger household economy and assessing impacts at the household level, the conceptual model deals with the problem of fungibility.

The conceptual model also helps to build the case for attribution by providing a plausible set of cause-and-effect relationships that link the microenterprise service to the impact. As described in Davis, "causal analysis in social research depends on assumptions about causal relationships" (Davis 1985, 66-67). The framework described in Sebstad et al. (1995) includes a discussion of the possible impact paths by which project interventions lead to positive changes at the household, enterprise, individual, and community levels. The framework identifies domains, which are broad categories of impact variables at each level.

The conceptual model allows microenterprise services to have both direct and indirect effects on the dependent variables in the causal system. The magnitudes of these effects are assumed to be conditional on the level of the treatment, which in this case would be the level or degree of microenterprise services received. The level of the treatment can be measured in different ways, including as the length of time that services were received, the number of services received, the monetary value of services received, or some combination of these variables.

In addition, impact levels are assumed to be affected by factors that act to moderate the cause-and-effect relationship between program services and impacts. Examples of such moderating factors include changes in the macro economy, market conditions in the sector and subsector, location of the enterprise, gender of the entrepreneur, household demographics, and the cultural/ethnic background of the household. Some of these moderating factors influence all households equally, so that they create the context for interpreting the results. Some of these factors differ across households, so that they need to be included as moderating (explanatory) variables in the analysis. The role of the moderating variables is discussed in more detail in section III below.

The remainder of this paper consists of four sections. The next section (section II) discusses the issues related to impact evaluation and describes how panel data and a non-experimental design help to address some of the challenges posed by selection bias. It closes with an overview of the key design features of the CIA. The third section focuses on the methods used in collecting and analyzing the panel survey data. This is followed by section IV, which describes the methods used in collecting and analyzing the case study data. Section V concludes the paper with a discussion of some of the advantages and disadvantages of the research strategy.

II. SELECTION OF THE RESEARCH DESIGN

A. The Impact Problem⁵

The basic challenge in impact assessment is to determine the effect of an intervention, or treatment, on an outcome variable. In the case of the AIMS core impact assessments, the intervention is the microenterprise service, which was credit in all three studies, plus savings services in India, and business training in Zimbabwe. The outcome variables are the variables in the core impact hypotheses listed earlier (e.g., household income, food expenditures, enterprise revenue). Moffitt provides this description of the impact problem:

“Suppose that we wish to evaluate the effect of a particular intervention (i.e., a treatment) on individual levels of some outcome variable. Let Y be the outcome variable and make the following definitions:

Y_{it}^* = level of outcome variable for individual i at time t if he or she has not received the treatment

Y_{it}^{**} = level of outcome variable for same individual i at same time t if he or she has received the treatment at some prior date.

The difference between these two quantities is the effect of the treatment, denoted τ :

$$Y_{it}^{**} = Y_{it}^* + \tau$$

or

$$\tau = Y_{it}^{**} - Y_{it}^*$$

The aim of the evaluation is to obtain an estimate of the value of τ , the treatment effect.” (Moffitt 1991, 292-293)

Impact evaluation seeks to measure the difference in outcome between an individual who received treatment and what the outcome *would have been* for the same individual, if he or she *had not received* the treatment. Obviously, the latter is an unobservable counterfactual event. The only practical alternative is to compare the outcomes for individuals who receive treatment with the outcomes for individuals who do not receive treatment. This is one of the fundamental problems in impact evaluation and the source of the associated selection bias problem.

B. Selection Bias in Impact Assessment

Selection bias exists if there are differences between those individuals who receive treatment and those who do not, and if these differences lead to incorrect measurement of the treatment effect. Selection bias can arise under situations in which an individual is free to choose (or elect) to receive treatment. In the case of the CIA, the program participants (borrowers and/or savers) all

⁵ The discussion and notation in this section are adapted from Moffitt (1991).

elected to receive the “treatment” by becoming clients and participating in the program. Similarly, selection bias can arise when program administrators choose who will be able to participate in a program and who will not. In selecting participants, administrators may logically choose to select only those applicants with specific characteristics that will make them more likely to succeed as clients.

It is plausible that individuals who participate in microenterprise programs differ, on average, from those who do not participate. In the context of microenterprise programs, many of the differences between participants and non-participants may relate to entrepreneurial ability: the ability to find and use information about the program, willingness to take a risk and try the program, and the ability to comprehend and realize the potential benefits of the program. Differences in the inherent entrepreneurial abilities of participants and non-participants may contribute to differences in an outcome variable, such as microenterprise revenue. These types of differences in outcomes should not be attributed to the program.

Another way to understand the problem that selection bias poses to the measurement of the treatment effect, or true impact, of a program is to consider the separate components of the treatment effect:⁶

$$\begin{array}{rclclcl} \text{Treatment} & = & \text{Measured outcome} & - & \text{Measured outcome} & - & \text{Outcome differences} \\ \text{effect (“true”} & & \text{for participants} & & \text{for non-participants} & & \text{due to selection bias-} \\ \text{impact)} & & & & & & \text{related differences} \\ & & & & & & \text{between participants} \\ & & & & & & \text{and non-participants} \end{array}$$

When there are selection bias-related differences between the participants and non-participants, and when these differences have a positive effect on the outcome, then the measured difference in outcome between the participants and the non-participants overestimates the true impact of participation. This would be the case for the CIA studies if, for example, the participants have greater entrepreneurial ability that would have led them to have a better outcome than the non-participants, even in the absence of program participation.

C. Experimental and Non-Experimental Designs

The preferred method for dealing with selection bias in impact assessment is to follow what is known as an “experimental” or “randomized” research design. In an experimental design, applicants to the program who satisfy the program eligibility requirements are randomly assigned to receive or not receive program services. Using this procedure, it is reasonable to assume that any differences between participants and non-participants are random, and that outcomes for the treatment group can be compared to outcomes for the control group with a known level of statistical confidence. Therefore, in the equation for the components of the treatment effect

⁶ To simplify the discussion, it is assumed here that the participants possess selection bias-related *advantages* leading to a positive effect on the outcome variables. Therefore the outcome differences due to selection bias must be subtracted from the measured outcomes to derive the true treatment effect. However, the implications on measurement of the treatment effect are conceptually the same for selection bias-related *disadvantages* leading to a negative effect. In that case, the last term on the right-hand-side would be added instead of subtracted.

given above, the last term would be zero, and the measured difference in the outcome would be equal to the actual treatment effect.

While an experimental design is optimal for the measurement of program impact, it is unusual for the administrators of an on-going program to be willing to randomly assign program participants.⁷ In the case of the CIA, it was considered inappropriate for microenterprise service providers to adopt a random assignment procedure, both for ethical and public relations reasons. In addition, it would have been logistically difficult to conduct survey interviews with clients in the limited number of days between their initial entry into the program and the receipt of their first loan. When an experimental design is infeasible, as was the case with the CIA studies, the alternative is to follow a non-experimental design, also known as a “quasi-experimental” or “non-randomized” design.

In a non-experimental design, the outcome variable is measured for the treatment group and for a constructed control group. The respondents in the control group do not receive the treatment but they are considered “similar” to the treatment group in critical ways that affect outcomes. Rossi and Freeman (1989) describe several different approaches for selecting control groups. The most commonly used method for constructing a control group is to select respondents that share critical characteristics with the treatment group, then to control statistically for differences in other variables that are expected to affect outcomes (Rossi and Freeman 1989, 328ff.). For example, if the treatment and control groups have different proportions of males and females, and if gender is expected to affect the outcome variable, then the gender differences between the treatment and control groups should be controlled for when statistically estimating the treatment effect.

In summary, while an experimental design is the preferred way to address the problem of selection bias, it is often infeasible. In the absence of an experimental design, a non-experimental design is the next best alternative. With a non-experimental design, however, additional measures are needed to reduce the distortion that selection bias can create in the measurement of the treatment effect.

D. Use of Panel Data in Non-Experimental Designs

Moffitt (1991) describes three general approaches to addressing the selection bias problem inherent in program evaluation with non-experimental data. He indicates that one of these approaches, the use of longitudinal data, has important advantages: “the availability of longitudinal data can eliminate the selectivity bias that would be present in only a single cross section of data” (Moffitt 1991, 298). Longitudinal data sets contain two or more measures over time. With cross-sectional data, respondents are only interviewed at a single point in time. When cross-sectional data are the only type of data available, then an attempt is usually made to reduce selection bias through econometric estimation of a fixed effects model.⁸

⁷ In the United States, experimental designs have been used to evaluate pilot programs that have limited funding relative to the number of qualified applicants. For an example of an experimental design study related to microenterprises, see Benus, Wood, and Grover (1994).

⁸ Many of the more recent impact assessments of microenterprise programs have been based on cross-sectional data sets. For example, the impact assessments described in Pitt and Khandker (1996), Khandker (1998), Lapaar et al.

Panel data, a specific type of longitudinal data, result from interviewing the same respondents on the same variables, measured at more than one point in time. There are two types of panel data that are important in impact assessment: 1) prospective panel data, drawn from multiple surveys that occur at different points in time, interviewing the same respondents and measuring the same outcome variables each time and 2) retrospective panel data, drawn from a survey that occurs only once, with respondents asked about outcome variables in two or more time periods (Menard 1991, 4-5).⁹

The use of panel data in a non-experimental design can be represented in terms of a two-by-two matrix in which the top row contains the outcome measurements for the treatment group and the bottom row contains the outcome measurements for the control group. In the case of panel data collected at two points in time, the first column contains the earlier (baseline) data and the second column contains the later (second round) data. In the matrix shown in figure 2, the letter “A” represents the measurement on the outcome (impact) variable for the treatment group during the baseline period, and B represents the measurement on the same outcome variable in the second time period. Similarly, C and D represent the measurements on the outcome variable for the control group at the baseline and second-round periods, respectively.

	Baseline Period	Second Round
Treatment Group	A	B
Control Group	C	D

Figure 2: Non-Experimental Panel Design

This matrix representation can be used to illustrate two of the common types of research designs for impact assessments. If data were collected to fill only the second column (cells B and D), this would correspond to a non-experimental design based on cross-sectional data. As discussed above, cross-sectional data do not provide the baseline information that is useful in dealing with selection bias. If data were collected to fill only the top row (cells A and B), this would correspond to a “before-and-after” study of participants only. The failure to include a control group in a before-and-after study can lead to serious questions about whether the measured changes should be attributed to the program or are, instead, at least partially the result of external, non-program influences.

In order to understand the advantages of a non-experimental panel design, consider the situation illustrated in figure 3. When the second round of data are collected (at time t), the treatment

(1995), and Zeller et al. (1996) rely on cross-sectional data. The impact assessments reported in Hulme and Mosley (1996) rely on either cross-sectional data or, in some cases, retrospective panel designs.

⁹ A third possible type of panel design is a repeated cross-sectional design, in which the survey is repeated two or more times, but with different respondents each time. This is less useful for impact evaluation because it cannot be used to establish causal order (Menard 1991, 27).

group has an average income of \$25,000. The average income of the control group is much lower, at \$10,000. Thus, the measured income gap between the treatment and control groups is \$15,000. This corresponds to the difference between B and D in figure 2 and represents the type of outcome data available through a cross-sectional design. However, this income difference (gap) of \$15,000 would overestimate the true treatment effect, because it includes outcome differences that are due to selection bias-related differences between the treatment and control groups.

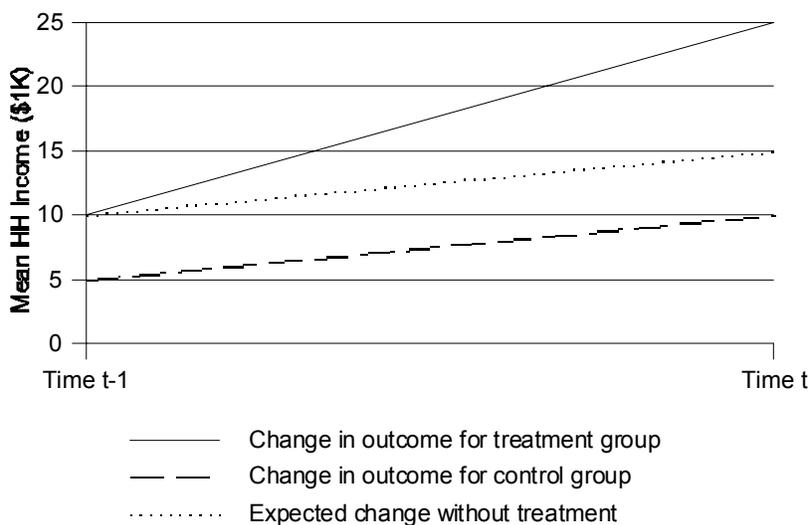


Figure 3: Example of Changes in Household Income in Non-Experimental Panel Data

The selection bias-related differences between the treatment and control groups are reflected in the differences already existing in the baseline period. At the time of the initial measurement (time t-1), the treatment group has an average household income of \$10,000, while the control group has a lower average income of \$5,000. This initial income gap of \$5,000 is related to the same observable and unobservable characteristics, such as entrepreneurial ability, that play a role in determining who participates in a program and who does not participate. This initial income gap explains at least some of the income gap in the later time period (time t).

A simple before-and-after study that included data for participants only would also overestimate the treatment effect. In such a study, the measured change in the participant group’s average income from \$10,000 in time t-1 to \$25,000 in time t reflects, at least partially, increases in income that are due to external factors, such as general improvements in the macroeconomic climate. The fact that the control group also experienced an increase in income between the two time periods suggests that at least some of the increase in income for the treatment group should be attributed to external factors other than program participation.

Given the initial differences between the treatment and control groups, a more accurate measure of the treatment effect (") is the difference between 1) the change in the outcome variable for the treatment group (B - A) and 2) the change in the outcome variable for the control group (D - C):

$$" = (\text{change in outcome for treatment group}) - (\text{change in outcome for control group})$$

$$\begin{aligned} " &= (B - A) - (D - C) \\ " &= (\$25,000 - \$10,000) - (\$10,000 - \$5,000) \\ " &= \$15,000 - \$5,000 = \$10,000 \end{aligned}$$

This measure accounts for both the fixed effects of selection bias differences between the treatment and control groups and the exogenous effects on outcomes that are unrelated to program participation.

Finally, it should be noted that this approach for measuring the treatment effect assumes that, in the absence of participation in the microenterprise program, the treatment and control groups would have similar rates of change in the outcome variables. In terms of figure 3 above, the assumption is that, in the absence of program participation, the slopes of the two lines would be equal, and the expected change for the “treatment” group would correspond to the dotted line. With additional data points in the longitudinal data set, it would be possible to test the assumption of equal slopes (Moffitt 1991, 302). However, with data for only two points in time, the assumption of equal rates of change in the absence of treatment must remain untested.

E. Key Design Features of the CIA

The purpose of the core impact assessments is to draw strong, plausible inferences about the impacts of microenterprise services on the clients, their enterprises, and their households. Designing an impact assessment is a challenging task, and there are several critical decision points. Unfortunately, there is no infallible approach that is guaranteed to reveal the irrefutable “truth” about the impacts of microenterprise services. Instead, there are alternative methods, each with advantages and disadvantages:

Regardless of the chosen design and the elaborateness of comparisons, however, some uncertainty about the size of treatment effects will always remain. It is impossible to rule out completely all threats to validity. Ultimately, researchers must rely on accumulating evidence across multiple designs and the corresponding multiple estimates of effects. (Reichardt and Mark 1998, 224).

The approach taken in the CIA was to combine quantitative and qualitative methods to reach a new level of understanding about the clients of microenterprise programs, the impacts of microenterprise services, and the sizes of these impacts, while recognizing the limitations of each method. The key features of the research design common to the three studies can be summarized as follows:

- The research was based on a conceptual framework of the microenterprise as embedded in the household economic portfolio.
- The data collection and analysis were designed to test a set of impact hypotheses at the enterprise, household, and individual levels.

- A mixed-method approach was followed, combining survey and case study data. The survey data provided information on the direction and size of impacts, while the case study data provided insights into the processes by which these impacts occur.
- The data were longitudinal, with two surveys administered over a two-year interval. The surveys were administered at the same time of year to control for seasonal differences. The same respondents were tracked over time, resulting in a panel data set. Two sets of case study interviews were administered over a one-year interval.
- A formal protocol was followed in collecting the case study data, resulting in the assembly of a case study database, which included extensive documentation of all interviews.
- The sample design for the survey was quasi-experimental, including both clients of the microenterprise program and non-clients with similar characteristics.
- The methods used to analyze the survey data included ANOVA tests, t-tests, chi-squared tests, gain score analysis, multiple linear regression, and analysis of covariance (ANCOVA).

The rest of this paper discusses the implementation of this research design. The methods for collecting and analyzing the quantitative data are described in detail in the next section. Section IV describes the qualitative component of the research.

III. COLLECTION AND ANALYSIS OF PANEL SURVEY DATA

A. The CIA Panel Data

The quantitative component of the CIA is based on a non-experimental research design utilizing panel data. More specifically, the CIA data correspond to a prospective panel design in which the survey occurs at two points in time. The same respondents were interviewed both times, and the same operational variables were measured each time. Another name for this research design is a *nonequivalent group design* (Cook and Campbell 1979).

Using the traditional notation to describe sampling designs in impact studies, the CIA panel data can be represented as follows:

$$\begin{array}{ccc} X & O & O \\ \hline & O & O \end{array}$$

where X represents initial receipt of microenterprise services, O represents the collection of data, and the dashed line is used to indicate that the treatment and control groups are nonequivalent. Program participants may have received one or more program loans at the time the baseline data were collected. In addition, the sizes of the loans received differed across

participants. Thus, the amount and degree of services received differed across the members of the treatment group.

Under this design, the initial data measurements represent a baseline, but they are not a “pre-test” *per se* because the baseline survey occurs after the initial treatment. This poses a potential problem. To the extent that the treatment has an effect on the baseline measurements, then the use of the baseline data to account for selection bias differences may also remove all or part of the treatment effect. In other words, the fact that the baseline data were not collected prior to the initial receipt of services means that the baseline data for the treatment group may already have been affected by the treatment. If this is true, then the resulting measures of the treatment effect will tend to underestimate both the positive and negative impacts of the microenterprise services.

In Peru and Zimbabwe, the survey data were collected in 1997 and 1999. The India survey data were collected in 1998 and 2000. In all three countries, the second survey rounds were repeated at the same time of the year as the baseline surveys. This was done in order to eliminate any distortions due to seasonal differences. A two-year interval between survey rounds was selected for several reasons: 1) one year was not considered a sufficient length of time to observe measurable impacts for certain outcome variables, such as asset accumulation; 2) two years was the longest interval that could be accommodated within the time period available for the AIMS Project; and 3) the budget available for the CIA would not cover three survey rounds over the two-year period (i.e., one survey per year).

Each study included a group of program participants and a constructed control group that shared similar gender, sector, and location characteristics with the participants. The treatment and control groups in all three countries were selected randomly. Approximately 700 households were surveyed in the Peru and Zimbabwe baselines, while 900 households were included in the India baseline.¹⁰ The sizes of baseline samples were selected to allow for panel attrition while still resulting in panel data sets that were large enough to allow for sufficient statistical power in testing the hypotheses.

B. Measurement of Core Impact Variables

The core impact variables at the household, enterprise, and individual levels are listed in tables 1-3 at the end of this document. For each variable, the tables indicate the number of the corresponding hypothesis and the specific measure that was used in each country. Also listed are any moderating variables that were consistently incorporated into the analysis in all three countries. The tables indicate the type of measurement scale used for each impact variable: 1) nominal, 2) ordinal, 3) interval, and 4) ratio (Godsey 1996, 15-17; Singleton, Straits, and Straits 1993, 110-118).

Nominal variables are also referred to as “categorical,” “qualitative,” or “dummy” variables. Nominal variables are used to identify categories and have no intrinsic numeric content. Some

¹⁰ The results of the baseline surveys in India, Peru, and Zimbabwe, are described in Chen and Snodgrass (1999), Dunn (1999), and Barnes and Keogh (1999), respectively. These reports also provide a detailed description of the procedures for selecting the treatment and control groups.

common examples of nominal variables include gender (male/female), participation status (client/non-client), location (rural/urban), and marital status (married/single/divorced/widowed). In order to be useful in quantitative analysis, the set of categories for each nominal variable should be mutually exclusive and exhaustive. In other words, each observation should fit in one, and only one, of the available categories.

Ordinal variables imply some type of ranking relationship. The different responses associated with an ordinal variable can be ranged in order of preference (better to worse) but there is no indication of the degree or absolute quantity by which one level is preferred to the other. Some examples of ordinal measurements include various types of opinion measures (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree) and variables that involve the ranking of alternatives (e.g., most important, second most important, and so on, down to least important).

Interval and ratio measures are related to cardinal numbers and are intrinsically numeric. Both measures are based on some standard measurement unit, or metric, so that it is possible to determine the distance between responses. With an interval measure, however, the starting point for the metric is arbitrary. The classic example of an interval measure is the Fahrenheit temperature scale, which has an arbitrarily selected zero point. Another example of an interval scale results from questions that ask the respondent to answer on “a scale from 0 to 20”. The starting point is arbitrary, because the exact same question could be answered on a scale from 20 to 40. While variables measured on an interval scale can be manipulated with addition and subtraction, they are not appropriate for the calculation of percentages, because the starting point for the metric can affect the result.

Variables measured on a ratio scale are numeric and have a non-arbitrary starting point. Examples of ratio variables would include monetary measures, measures of length of time, and counts of the number of some type of object. Ratio scales are convenient, because they can be analyzed using all types of mathematical operators (e.g., addition, subtraction, percentage). As can be seen in tables 1-3, the majority of the impact variables were measured on ratio scales. However, one each of the household-level and enterprise-level variables, and all of the individual-level variables, were measured on nominal scales.

C. Sample Design and Data Collection

1. Sample Selection

In each of the three studies, a two-stage sampling approach was followed. The first stage was to choose geographic regions that were representative of the program’s overall client base while containing a large concentration of clients within close proximity of each other. This sampling approach improved the cost effectiveness of the survey, primarily through the cost savings derived from limiting the geographic coverage needed in constructing the non-client sample frame. In addition, confining the survey to a limited number of areas generated savings in enumerator salaries, transportation costs, and other logistical costs.

The second stage of the sampling approach consisted of the selection of the client and non-client households. Client households were randomly selected from updated client lists provided by the

microenterprise support programs. Once the client households were selected, then comparable non-client households were randomly selected from within the same neighborhoods as the clients in the sample. In India and Peru, the non-client households were randomly selected from a constructed sample frame; in Zimbabwe, a random walk procedure was used to select the non-client households. In all of the sites, the non-clients were screened on the basis of their similarities to the clients and their eligibility for program participation.

2. Participation Categories

In order to analyze the data, it was necessary to first determine which observations would be included in the client (treatment) sample and which observations would be included in the non-client (control) sample. The approach was to differentiate between the client and non-client samples according to the participation status of households at the time of the baseline survey. In addition, each study accorded special attention to specific subgroups of the client sample that were defined in terms of the ways they had participated in the program.

The **client sample** (a.k.a. the “treatment” or “participant” group) included all households who were classified as program clients at the time of the baseline survey. Specifically, the client sample included 1) all households who were clients of the programs in both the baseline and second-round periods and 2) all households who were clients during the baseline, but who were no longer clients at the time of the second-round survey. The **non-client sample** (a.k.a. the “control” or “non-participant” group) consisted of all households who had never received services from the program being studied and who met other comparability and eligibility criteria.

There is obvious justification for including in the client sample all households who were clients of the microenterprise program at the time of both the baseline and second-round surveys. In addition, households who were clients during the baseline, but who were no longer clients at the time of the second-round survey, were also included in the client sample because they had received the “treatment” and could be expected to demonstrate a “treatment effect.” Their inclusion in the client sample was based on the argument that a true measure of a program’s impact should include impacts on those who are no longer receiving services.

Because the SEWA Bank program offered both credit and savings services, the India study included two treatment groups: 1) savers only and 2) borrowers who were also savers. The **savers** were SEWA members who had active savings accounts with SEWA Bank but did not have an outstanding loan at the time of the baseline survey. However, the savers may have borrowed from SEWA Bank at other times. The **borrowers** were SEWA members who had an outstanding loan from SEWA Bank at the time of the baseline survey. All borrowers also held savings accounts in SEWA Bank. The impact analysis in the India study consisted of a three-way comparison between savers, borrowers, and non-clients. This provided information on the differential impacts of different types of financial services.

In the Peru study, special attention was given to those households who were non-clients at the time of the baseline survey, but who received their first microenterprise loan in the interim period between the baseline and the second round. In this **new entrant** group, the households had received at least some program services and could potentially be expected to experience

impacts. However, since they received their initial program services relatively soon before the second-round survey, the level of impacts for the new entrants was expected to be lower than for households in the client group. The new entrant group was analyzed separately to provide additional insights into the timing of impacts and the size of the treatment effect relative to the size of the selection bias.

By comparing the impacts on the new entrants to the impacts on the main treatment group, it is possible to gain some information about whether selection bias is a major problem in the data. If the measured treatment effect for the new entrant group (in comparison to the control group) is smaller than the treatment effect for the client group (again, in comparison to the control group), then that can be interpreted as providing evidence that the analysis is not seriously biased by selection differences. The new entrant group and the main treatment group can be expected to be similar in terms of the factors related to selection bias. Therefore, if they differ in terms of the magnitudes of impacts, then those differences are more likely to be attributable to different levels of program participation than to selection bias.

In the Zimbabwe study, special attention was given to the level of continued program participation after the baseline. The client sample was separated into the **departing clients**, those who took no additional program loans after the baseline, and the **continuing clients**, those who took one or more additional program loans after the baseline. To supplement the standard comparison of clients to non-clients, the statistical analysis included a three-way comparison between departing clients, continuing clients, and non-clients. Moreover, for certain impact variables, additional distinctions were made based the level of program participation prior to the baseline survey. The client sample was separated into these subgroups based on the assumption that different levels of program participation would be associated with different levels of program impacts.

3. Tracking Respondents and Enterprises

Between the first and second rounds of the survey, certain types of changes affected the ability to track the units of analysis over time. These changes occurred at the household, enterprise, and individual levels. At the household level, a critical change was the change in program participation status, which was discussed in the section immediately above.¹¹ At the enterprise and individual levels, respectively, the closure of the primary enterprise and the death of the primary respondent made it impossible to compile critical longitudinal data at these specific levels of analysis.

For the analysis of the data, the following tracking rules were applied:

- 1) If the household could not be located for the second round, or refused to participate in the second-round survey, then that observation was removed from the longitudinal data set.¹²

¹¹ In addition, in Zimbabwe there were a few cases in which a household joined with one or more members of another household between the baseline and second-round surveys.

¹² Rates of panel attrition and the analysis of panel attrition are reported in each of the three studies. The analysis of panel attrition is also discussed below (under “Data Analysis”).

- 2) If the primary enterprise from the baseline had closed or otherwise did not exist at the time of the second-round survey, then that observation was retained in the data sets for the analysis of household-level and individual-level impacts, but it was removed from the data set for the analysis of impacts on the primary enterprise. A separate analysis was conducted on the reasons for closure and, if there were enough cases of new primary enterprises, these were compared to the closed primary enterprises.
- 3) The enterprise-level impact hypotheses were analyzed twice: once for primary enterprises only (see rule #2 above) and once for the primary plus up to two additional enterprises associated with the household. The aggregated analysis included only hypotheses E-1 and E-3 (note that the aggregated value of all enterprise fixed assets was tested in hypothesis H-3c).
- 4) If the primary respondent in the second round differed from the primary respondent in the baseline (due to death, illness, or other unavailability), then that observation was removed from the data set used in the analysis of the individual-level impact hypotheses.¹³

D. Data Analysis

The survey data were analyzed using several complementary approaches. In order to provide information on the relative changes in the outcome variables between the two rounds of the survey, paired t-tests and gain score analysis were used. The results of the paired t-tests and gain score analysis were treated as part of the descriptive analysis, rather than as part of the impact analysis *per se*. In other words, these results provided information about the size and direction of changes in the outcome variables for all of the comparison groups, which is important background information for interpreting the impact results. However, conclusions about impacts were derived primarily from the results of the ANCOVA procedure, which estimated both the impact of the treatment as well as the influence of other explanatory, or moderating, variables.

1. Measuring Initial Differences in the Baseline Data

The methods for analyzing the baseline data included cross tabulation and testing for statistically significant differences between mean values. The cross tabulations highlighted initial differences in the mean values of the hypothesized impact variables between important subgroups, such as between clients and non-clients and between old clients and new clients. In addition, cross tabulations were created for other important subgroups as defined by, for example, enterprise sector and gender of the entrepreneur.

In the cases where the differences between the mean values of the impact variables were of a sufficient magnitude to be of some interest, then tests for the statistical significance of these differences were conducted. These tests included the following:

- t-tests, for comparing two means measured numerically;

¹³ For the India study, if the primary respondent from the baseline was unavailable to participate in the second round, then the entire household and all of its associated enterprises were eliminated from the longitudinal data set.

- ANOVA tests, for comparing three or more means measured numerically; and
- chi-squared tests, for comparing differences in the distribution of categorical data.

This initial analysis was helpful in indicating whether any of the comparison groups were starting with significantly different levels of the impact variable and identifying variables, such as sector and gender, that might be associated with these differences. This information was used to indicate the possible existence of selection bias in the sample. It was also useful in identifying moderating variables to be used in the impact analysis. The analysis of the baseline data provided a great deal of descriptive information about the clients of the microenterprise support programs. This descriptive information about clients was useful to program managers.

2. Measuring Changes Between the Two Survey Rounds

a. Paired t-tests

In the Peru and India studies, paired t-tests were used to determine whether any of the comparison groups experienced significant changes in the levels of the outcome variables between the two rounds of the survey. For example, in order to determine whether the level of food expenditures had changed significantly for the treatment group between 1997 and 1999, a paired t-test was performed in which the levels of food expenditures for households in the treatment group in 1997 were compared to the levels of food expenditures for the same households in 1999.

Sets of paired t-tests were calculated for each of the outcome variables and each of the comparison groups. When the results of a paired t-test were significant, that indicated that there was a statistically significant change in the mean value of the outcome variable between the two rounds of the survey. It did not indicate, however, whether the change in the outcome variable was due to program services or to some other influence.

b. Gain score analysis

Gain score analysis was used to determine whether the impact variables followed significantly different trends over time for the different comparison groups. The basic idea in gain score analysis is to compare the treatment and control groups in terms of the average gain in the impact variable experienced by each group between the first and second rounds of the survey. In comparing the changes in outcomes for program participants to the changes in outcomes of non-participants, gain score analysis is an intuitively straightforward data analysis approach. By taking account of the fact that each group may have a different starting point in the baseline survey, gain score analysis addresses some of the main effects of initial selection differences (Reichardt and Mark 1998, 216).

The conceptual underpinnings of gain score analysis have already been presented in section II. In terms of the matrix in figure 2, gain score analysis is based on comparing the average change in the impact variables for the treatment group to the average change in the impact variables for the control group. This corresponds to the calculation of the treatment effect ("") as the

difference between (B - A) and (D - C).¹⁴ If the mean value of the change experienced by the treatment group (B - A) is greater than the mean value of the change experienced by the control group (D - C), and the difference between the means is statistically significant, then that provides empirical evidence to reject the null hypothesis that the means are equal.¹⁵

c. Statistical and “real-world” significance

The calculated mean gain scores are statistical estimates that have variances associated with them. Because they are estimates, any differences between the mean values for participants and non-participants were tested for statistical significance before concluding that the two groups experienced different rates of change between the two survey rounds. The differences between the mean values for participants and non-participants were tested for statistical significance using a t-test.¹⁶ For consistency with standard practice and uniformity across the three studies, the levels of statistical significance that were reported were .01, .05, and .10.

Once statistical significance had been established, then the real-world significance of the difference was also considered: Is the difference meaningful in a real sense? In other words, while a difference between \$200 and \$201 in monthly food expenditures may be statistically significant, it is probably not meaningful in dietary terms. A useful way to explore the real-world significance of a result is to construct confidence intervals, which show the likely range of a variable and the degree of uncertainty about its true size (Reichardt and Gollob 1997).

d. Analyzing Attrition

Two types of attrition were analyzed: panel attrition and program attrition. Panel attrition refers to the loss of survey respondents between the two rounds of the survey. This was analyzed both to determine the level of attrition and to determine whether the loss of respondents changed the representativeness of the sample. According to Burgess (1989), it is considered reasonable in panel studies to retain 80 to 90 per cent of initial respondents. The standard procedure for dealing with panel attrition is to use the baseline data to determine whether there were significant differences between the “panel leavers” and the “panel stayers.”

Two types of analysis were conducted: 1) testing whether the proportion of the sample in important categories had changed significantly over time and 2) testing whether the panel leavers within a particular category were significantly different from the stayers in terms of their initial values on particular variables (or strength of relationships between variables). The variables that were analyzed included the following:

¹⁴ There is some debate surrounding the use of raw change scores as opposed to residual gains or lagged variables. Given the research design for the CIA and the predominance of economic impact variables in the data, the use of raw change scores appears to be justified (Menard 1991, 46-47; Liker, Agustyniak, and Duncan 1985).

¹⁵ For information on how to calculate gain scores, see the appendix.

¹⁶ When testing for statistically significant differences between three or more means, an ANOVA test should be performed first to determine if there are statistically significant differences between any of the means. If the results of the ANOVA test are statistically significant, then a series of t-tests can be performed on different pairs of means. Pairs of means can also be tested using a Tukey test.

- participation status (client/non-client),
- gender (for the Peru and Zimbabwe data only),
- sector (commercial/service/industrial),
- location (site-specific survey areas), and
- employment status groups (for the India data only).

In addition, for those who were clients in the baseline, the loan histories for panel leavers and panel stayers were compared in terms of 1) number of loans or total amount of loans to date and 2) loan repayment record since baseline. The results and implications of the panel attrition analysis are reported in each of the three final reports.

Program attrition refers to the loss of clients from the microenterprise support program. It was analyzed both to generate information for program managers and to determine its relationship to panel attrition. For those who were program participants in the baseline, but who were no longer program participants at the time of the second round, these characteristics were examined:

- gender (for the Peru and Zimbabwe data only),
- sector (commercial/service/industrial),
- location (relative to site-specific survey areas),
- employment status groups (for the India data only), and
- whether primary enterprise was still in operation.

In addition, the loan histories for program leavers and program stayers were compared in terms of 1) number of loans or total amount of loans to date and 2) loan repayment record since baseline. This analysis provided useful information on program attrition to the managers of the microenterprise programs.

In each of the three studies, the data were analyzed to determine whether a significant number of households had received similar services from some other microfinance program(s). In the case of the Peru and Zimbabwe studies, the non-client households were screened into the baseline on the basis of not having received any other type of program or formal credit. However, some households began to receive these services in the interim period between the first and second rounds of the survey. In Peru, where a significant number of respondents received similar microenterprise credit from alternative sources, the possibility of alternative microenterprise credit was brought into the analysis.

3. Controlling for Moderating Variables

For each of the impact variables, there were key moderating variables that were believed to affect the relationship between program participation and the change in the impact variable. That is, the nature of the relationship between program participation and impact was expected to be different for different levels of the moderating variables. Several of the specific moderating variables included in the analysis are listed in tables 1-3. Additional moderating variables were selected on a site-specific basis.

The different types of moderating variables, and the role they play in affecting impacts, are discussed in previous AIMS papers (Sebstad et al. 1995; AIMS Team 1996). For example, the sector of the enterprise (commercial, service, production) is considered to be an important moderating variable in that it is assumed to affect the link between program participation and changes in enterprise-level impact variables. Similarly, client gender is considered an important moderating variable for certain impacts at the enterprise, household, and individual levels.

There are at least three reasons for including moderating variables in the analysis. First, if the comparison groups differ significantly in the distribution of these moderating variables, then at least some of the measured differences in impact may be due to differences in the moderating variables rather than to program participation alone. In this case, it is necessary to make statistical adjustments in order to control for differences between the comparison groups in the moderating variables. Including the moderating variables in the impact analysis in this way helps to control for initial selection differences. Second, by including the moderating variables, it may be possible to determine how the size of the outcome variable differs according to the level of the moderating variables. Third, the moderating variables may help to statistically explain the variation in outcomes, thus providing a more powerful analysis of impact.

For the gain score analysis, some of the moderating variables were used to subdivide the data for the treatment and control groups (Rossi and Freeman 1989, 328-331). For example, the data on enterprise revenues were subdivided by sector in order to compare changes in revenues for treatment and control enterprises in the same sector. A second subdivision might be used to control for the effects of another key moderating variable, such as location of the enterprise.

This subdivision approach to controlling for differences between the treatment and control groups is only practical for one or two levels of moderating variables. After that, the format for presenting and interpreting the results becomes too cumbersome. The ANCOVA approach, described in the next section, statistically controls for multiple moderating variables simultaneously.¹⁷ With either method, there is a limit on the number of moderating variables that can be included in the analysis, since the number of observations in each subgroup may become too small to derive statistically significant results.

4. Analysis of Impact with ANCOVA

The analysis of covariance (ANCOVA) procedure was the central approach used to analyze the panel data and test the hypotheses about the impacts of microenterprise services. While the ANCOVA procedure is more complex than gain score analysis, ANCOVA has the advantage of statistically controlling for multiple differences between the treatment and control groups, differences that may have a moderating effect on the relationship between program participation and changes in the impact variables.

In effect, the ANCOVA procedure provides a statistical “matching” of individuals in the treatment and control groups who have the same baseline measures on the impact variable and

¹⁷ Note, however, that the subdivision and simultaneous approaches are both possible in gain score analysis and the ANCOVA approach.

on the moderating variables.¹⁸ It then compares these matched observations to determine whether there are any consistent differences between the treatment and control groups in terms of second-round outcome values. In other words, given similar measures on the baseline values of the impact variable and the moderating variables, the ANCOVA procedure looks for systematic differences in second-round outcomes. It “statistically matches individuals in the two treatment groups on their pretest scores, and uses the average difference between the matched groups on posttest to estimate the treatment effect” (Reichardt and Mark 1998, 217-218).

Essentially what ANCOVA does is to allow separate, parallel regression lines to be fitted through the data for the treatment group and the data for the control group.¹⁹ The regression lines estimate the second-round measure of the outcome variable, given the measure of the outcome variable in the baseline. To the extent that participants have systematically higher second-round measures for a given baseline measure, then the regression line for the treatment group will have a higher intercept than the regression line for the control group. In ANCOVA, the distance between the two regression lines is the estimate of the treatment effect. The ANCOVA procedure can be implemented with the SPSS software.²⁰

The ANCOVA approach can be used to control for a variety of differences between the treatment and control groups. In other words, specific moderating variables, such as gender of the entrepreneur and sector of the enterprise, can be brought into the ANCOVA model as additional “covariates.” As with the baseline measures on the outcome variable, ANCOVA statistically controls for differences in moderating variables by matching observations with the same or similar levels. In addition to the variables listed in tables 1-3, any other moderating variables that can explain a substantial part of the variation between individuals within groups should be included in the ANCOVA analysis, since this helps to uncover statistically significant impacts, even when the treatment effect is small relative to the variability related to the moderating variables.

As an example, suppose ANCOVA is being used to estimate the effect of microenterprise services on microenterprise revenue, with the sector of the enterprise and the gender of the entrepreneur being important moderating variables. In this case, the covariates would be 1) the level of enterprise revenue in the baseline data; 2) sector of the enterprise; and 3) gender of the

¹⁸ In the language of ANCOVA, the baseline measure of the impact variable and the moderating variables are the “covariates.”

¹⁹ The ANCOVA model can be represented algebraically as follows (Reichardt 1979, 153):

$$Y_{ij} = \mu + \alpha_i + \beta(X_{ij} - \bar{X}) + \epsilon_{ij}$$

$\epsilon_{ij} \sim \text{NID}(0, \sigma^2)$
 X is fixed (if random, then X is independent of ϵ_{ij})

The notation is used differently here than in the rest of this document. Here, i indexes the group (treatment or control), and j indexes the individual observation (e.g., household). The variable Y is the second-round measure, μ is the mean value of all the second-round observations, X is the baseline measure, and \bar{X} is the mean of the baseline measures. The estimate of the treatment effect is the difference between the estimate of α for the participant group and the estimate of α for the control group.

²⁰ With SPSS version 7.0 or higher, the ANCOVA procedure is found under the multivariate option of the general linear model (GLM).

entrepreneur. The ANCOVA procedure would estimate the effect of enterprise services on enterprise revenues (i.e., the treatment effect) as the difference in intercepts between two parallel (multiple) regression lines for which the values of the three covariates are held constant (i.e., the observations are statistically matched to have similar levels on all three covariates).

It is possible to distinguish between ANCOVA, gain score analysis, and the use of only second-round (i.e., cross-sectional) data in terms of the degree to which each procedure incorporates information on the baseline measures into the estimate of the treatment effect (Frison and Pocock 1992). Broadly speaking, all three of these methods are subtracting some weighted value of the baseline measure from the second-round measure: [second-round measure - weight*baseline measure]. In gain score analysis, the weight given to the baseline measure is 1, while an estimate based only on cross-sectional data is equivalent to using a weight of 0 (i.e., no incorporation of information on the baseline measure). The ANCOVA procedure selects the weight (β) that provides an optimal correction as determined by the regression.

5. Conclusions on the Panel Data Analysis

The panel data were analyzed using a variety of approaches, with each approach designed to yield different types of information. When the gain score results were consistent with the ANCOVA results, the conclusions of the quantitative analysis were considered to have greater strength and credibility. The strengths and weaknesses in the quantitative approach are summarized in the final section of this paper. This next section focuses on the collection and analysis of the case study data. By triangulating the results of several quantitative methods with the results of the qualitative methods described in the next section, the research attempted to create an even stronger basis for drawing conclusions about the impacts of microenterprise services on the clients of the three programs.

IV. COLLECTION AND ANALYSIS OF CASE STUDY DATA

In order to develop an in-depth understanding of program impact, the core impact assessments relied on a mixed-method approach. The quantitative approach, which included both the longitudinal survey and the use of clients' credit history data, provided valuable information on the "who, what, where, how many, and how much" of program impacts. The qualitative approach taken in the case studies was designed to augment the survey data by answering the "how" and "why" questions (Yin 1994). Thus, the case studies were designed to supplement the survey and were based on a case-within-survey design with pre-program and post-program impact measures.

A. Objectives of the Case Study Research

The overall objective of the case study research was to examine how and why changes occur as the result of program participation. Using the household economic portfolio model as a conceptual framework, the research focused on enterprise-, household-, and individual-level variables to address the following research question: *"How do microfinance services contribute to the observed changes within the household, its enterprises, and the individual client?"*

The specific objectives of the case study research were the following:

- 1) To understand how impacts occur in the context of resource management within the household economic portfolio;
- 2) To understand client perspectives on impact;
- 3) To test specific individual-level hypotheses;
- 4) To identify the sequence of events between program participation and impacts;
- 5) To test rival explanations for observed impacts; and
- 6) To interpret unexpected and unanticipated findings from the survey.

One of the most important, yet most difficult, tasks of an impact assessment is to provide convincing evidence that the measured changes, or impacts, can be attributed to the program being evaluated. While the survey results document changes in the impact variables between the two survey rounds and provide statistical evidence of any differences between clients and non-clients, the case studies were used to reconstruct the chain of events leading to those changes. By identifying the sequence of events leading from program participation to the measured impacts, the case study data complement and strengthen the survey results and can help to improve the case for attribution.

Another way that the case study research strengthens the overall case for attribution is by investigating, and possibly disproving, rival explanations for the observed impacts. If the survey results provide statistical evidence for impacts, and the case study results suggest that these impacts were not due to factors other than program participation, then greater confidence can be placed in the conclusion that the observed impacts were due to program participation.

The case study research was also used to investigate unexpected and unanticipated findings. Unexpected findings can occur when the variables identified in the hypotheses do not show the expected (or hypothesized) relationships or are inconsistent with the relationships identified in the conceptual model. Unanticipated findings can occur when important impacts are found which were not identified in the hypotheses or conceptual model. Case study research can probe for explanations of such unexpected and unanticipated findings.

B. Study Questions and Propositions

The study questions for the case study research were driven by the need to supplement the survey findings on the study's hypotheses, to better understand impact processes, and to examine rival explanations for the measured impacts. These study questions, which were derived from the same hypotheses that drove the survey research, constituted the foundation and focus for the collection and analysis of data in the case study research. Each study question has an accompanying research proposition, which describes a hypothesized pattern of events leading from program participation to changes in the impact variable. The patterns specified in the research propositions are consistent with the hypotheses and conceptual model.

In general, the overall question that motivated the case study research was the following:

What changes have occurred as a result of program participation; to whom and under what conditions did these changes occur; and, why did these changes occur?

As this overall question was too broad to be operationally useful, specific study questions and propositions related to the household economic portfolio and the impact hypotheses were developed. These questions and propositions are listed in table 4 at the end of the document. They served as a framework to guide the case study interviews.

The first three study questions examine changes in three specific areas: 1) the composition of the household economic portfolio; 2) financial and risk management behavior; and 3) intrahousehold control over resources and income. These categories were selected because they allowed for an analysis of interaction between a range of household (H-1, H-2, H-3, H-4, H-6), enterprise (E-1, E-2), and individual-level (I-1, I-3) hypotheses. Three of the remaining four study questions were more narrowly defined, focusing on changes in self-esteem (I-2), orientation toward the future (I-4), and the transaction relationships of the enterprise (E4a, E4b). The final question related to client perspectives on the impacts of program participation.

For all of these questions, the case study research examined WHAT changes occurred, TO WHOM they occurred, HOW (or the process by which) they occurred, and WHY they occurred. Thus, the data collection concentrated on gathering information relevant to the questions and propositions, uncovering the answers to the study questions, highlighting the processes by which changes occurred, and determining whether the patterns of change exhibited by the cases matched the patterns specified in the propositions.

C. Implementation Issues

1. Embedded Unit of Analysis

The household economic portfolio model provided the conceptual framework for the qualitative research. In this model, the microenterprise is embedded in the household economy. This naturally led to the selection of an embedded unit of analysis for the case studies, with the household serving as the main unit. Individuals within the households and the microenterprises associated with household members were treated as subunits that were embedded in the household unit (Yin 1994, 41-44). For consistency, the case studies used the same definitions for “household,” “microenterprise,” and “client” that were used in the survey research.

2. Selection of Cases and Replication Logic

The cases for the case study research were selected to fit the underlying assumptions and conditions that were relevant to the research questions. This type of selection, based on replication logic, strengthens the validity of the study by ensuring that the case is a relevant instance of the unit of analysis. Selection on the basis of replication logic stands in contrast to selection on the basis of a sampling logic, which was used for the survey.²¹

²¹ Sampling logic has the goal of gathering a preponderance of evidence so that the findings can be generalized to the underlying population with some known (specified) degree of confidence.

Only clients were included in the case studies because non-clients were not relevant to the overall research question: “What changes have occurred as a result of program participation; to whom and under what conditions did these changes occur; and why did these changes occur?” Because the purpose of the core impact assessments was to document and measure the impacts of program participation, the survey provided the data needed to make defensible comparisons between clients and non-clients and to measure how the clients and non-clients differed in terms of the impact variables. The case studies, on the other hand, were designed to shed light on the processes that led from program participation to impacts.

Multiple cases were selected in each country in order to gain literal and theoretical replication (Yin 1994). Literal replication occurs when more than one case points to similar results. The ability to replicate similar findings with multiple cases strengthens the credibility of the case study findings. Theoretical replication occurs when two cases point to contrasting results, but the differences between the cases are predictable and based on the underlying theoretical framework of the study. For example, different results were expected for poor and non-poor households. The cases were selected on the basis of several variables in order to provide for both literal and theoretical replication.

3. Selection Variables

The case study research included nine to twelve case study households,²² selected on the basis of level of program participation, household income or asset level (in Peru and India), and additional site-specific variables. Information from the baseline survey was used to identify which client households fit into each of the subgroups. Multiple cases were selected in each subgroup to provide literal replication, while the possibility of differences in impacts across the subgroups (e.g., poor versus non-poor; new versus long-term participants) allowed for theoretical replication.

Level of program participation was used as a selection criterion in all three studies, based on the premise that level of program participation affects the types of impacts as well as the intensity of these impacts. Participation was measured either in terms of number of loans or length of time in the program. The income or asset level of the household was also used as a selection variable in the India and Peru studies. According to the conceptual framework for the project, the income or asset level of the household is a determining factor in the size and composition of the household economic portfolio and can be expected to influence financial and risk management decisions. Other site-specific selection criteria included gender in the Peru study, trade group and employment history in the India study, and program in the Zimbabwe study, where there were two distinct lending programs.

4. Time Period for Analysis

For each round of interviews, the questions related to specific time boundaries and attempted to identify the patterns of change within those time periods. During the first round of interviews,

²² There were nine case study households in the Zimbabwe study, eleven in the Peru study, and twelve in the India study.

the attention was focused on changes that had occurred since the client joined the program, with particular emphasis on the changes that had occurred in the period immediately before and after joining the program. These first interviews reconstructed the history of the household and looked at the trend lines in the impact variables specified in the hypotheses, to understand how joining the microenterprise program fit into the economic history of the household. The second round of case study interviews were conducted with the same households that participated in the first round. These interviews focused on changes in key impact and moderating variables that occurred between the first and second rounds of the survey and sought explanations for those changes.

D. Analyzing the Case Study Database

1. Case Study Database

To organize the data from the case study research, a case study database was created for each household included in the study. This database holds the evidence that was used in the analysis of the case studies. The database consisted of the following components:

1. Summary table of credit data;
2. Summary table of survey data;
3. Narrative summary of history of credit transactions;
4. Hand written notes from each interview session, arranged by date of interview;
5. Word processed notes from each interview session, arranged by date of interview;
6. Transcriptions of tape recorded interviews; and
7. Narratives containing open-ended answers to study questions.

The seventh component, narrative answers to the study questions, is a critical feature of the case study database. Started in the field and finished when the transcriptions of the interviews were completed, the narrative answers to the study questions were written as open-ended answers based on the evidence collected prior to and during interviews. There is one narrative for each case that integrates and synthesizes information from the interviews in order to answer the seven study questions.

2. Data Analysis Through Pattern Matching

The case study data were analyzed using a variant of pattern matching called the program logic approach. With pattern matching, patterns in the empirical evidence are compared to the patterns predicted in the study propositions. If the patterns match, there is evidence in favor of the study propositions and the internal validity of the research is strengthened.

The program logic model is a type of pattern-matching analysis that includes an element of time-series analysis. A program logic model allows an examination of changes in the dependent variables over time. In this model, the program intervention is posited to lead to immediate and intermediate outcomes, which in turn lead to the final impacts. This approach, as described by Yin (1994), allows for the examination of the cause-and-effect relationships between independent and dependent variables over time.

The study questions and propositions that drove the case study research were derived from the core research hypotheses and the household economic portfolio model. Each of the study propositions posits a chain of events leading from program participation to a specified final impact. This made the program logic approach appropriate for analyzing the data gathered during the research. Analysis of the data in the case study database focused on comparing the empirical evidence from the interviews to the hypothesized cause-and-effect relationship between the receipt of program services and changes in a range of dependent variables (e.g., risk management, investment activities) over time.

3. Integration of Survey and Case Study Findings

The mixed-method approach of combining statistical hypothesis tests with in-depth qualitative data has the advantage of combining two complementary approaches within a single study. However, there is a challenge in integrating the survey findings with the case study findings. The final CIA reports integrated the survey and case study findings in different ways. One approach was to present the results of the statistical tests of the impact hypotheses and report the related material from the case study data alongside the statistical results. In this way, the case study material helped to explain the processes leading to the statistical findings. A second approach was to present the survey and case study data separately, using the case study data to develop descriptions of the household economic portfolios, financial practices, and risk behaviors. This approach provided rich descriptive information on the ways that clients interacted with the program. Both approaches for integrating the two types of results created a stronger case for attribution than would have been possible with a set of findings based solely on survey data.

V. SUMMARY AND CONCLUSION

The AIMS core impact assessments are serious attempts to rigorously measure and report the client-level impacts of microenterprise services. The findings of these studies provide many insights into the interaction between clients and programs. The findings also indicate, with a reasonable degree of credibility, that impacts are not always where one might expect to find them. In addition, the studies represent an advance in the development of methods for evaluating the impacts of microenterprise programs.

In planning the AIMS core impact assessments, considerable attention was placed on addressing the design challenges of fungibility, attribution, and selection bias. Selection bias, in particular, was considered a key threat to the internal validity of the CIA results. This paper has described the common features of the research strategy that was followed in the three studies. This final section summarizes the major strengths and weaknesses of the research approach and suggests some of the implications for the design of future studies.

A. Advantages of the Research Strategy

There were four major design features that contributed to the strength of the research strategy: the mixed-method approach, the implementation of a quasi-experimental design, the collection

of prospective panel data, and the selection of an appropriate conceptual framework. The advantage of each of these design features is summarized below.

- *Mixed-method approach.* The combination of survey and case study data yielded two types of information: 1) quantitative information estimating the size and direction of impacts and 2) qualitative information describing the processes by which impacts occur. The case study results complemented the survey results, yielding a much more informed view about how and why impacts occur and strengthening the case for attributing the measured impacts to the program services.
- *Quasi-experimental design.* By including in the survey both clients of microenterprise programs and comparable non-clients, it was possible to control for the influence of external factors unrelated to the program, such as macroeconomic conditions, that affect the outcomes for both clients and non-clients. The changes experienced by the control group over the period of the study provided some information on the kinds of changes the clients might have experienced anyway, even in the absence of program participation (i.e., the counterfactual).
- *Prospective panel data.* The availability of panel data allowed greater simplicity and transparency in the choice of analytical methods. Since panel data include information on the starting points of the impact variables, then impacts can be measured in terms of differences in the rates of change for the impact variables rather than as absolute differences in the impact variables in a single time period. This helps to eliminate some of the influence of selection bias on the results. In addition, data collected from two survey rounds (prospective data) are considered more reliable measures of the variables than retrospective data collected in a single survey.
- *Conceptual framework.* The conceptual and causal models underlying the CIA studies provided a logical framework for asserting that the observed impacts could be attributed to the program services received. The household economy approach of the conceptual framework allowed the research to address the problem of fungibility by measuring impacts at three levels.

These design features combined to yield strong, plausible inferences about the impacts of microenterprise services and to address some challenging design issues. Another way to discuss the advantages of the research strategy is to consider how the research design addresses the specific design challenges of fungibility, attribution, and selection bias.

- *Fungibility.* The research approach deals squarely with the problem of fungibility by widening the unit of analysis beyond the credit-supported microenterprise to include impacts on the entire household economy. The issue of fungibility of credit is resolved by broadening the unit of analysis and testing a number of impact hypotheses at three levels: the household, the enterprise, and the individual.
- *Attribution.* The research approach addresses the problem of attribution in several ways. First, the study is based on a conceptual model of the household economic portfolio that

provides a plausible link between the receipt of microenterprise services and the hypothesized impacts. Second, the study relies on a mixed-method approach that uses carefully collected and analyzed case study data to examine the counterfactual and to test for the existence of rival hypotheses. Finally, the use of a control group in the longitudinal survey helps to assure that any changes in the impact variables due to changes in the economic environment are not incorrectly attributed to the microfinance program.

- *Selection bias.* The use of a quasi-experimental design combined with panel data permitted the removal of some, but not all, of the influence of selection bias on the results. With the ANCOVA approach, it was possible to control for the starting values of the impact variables and to statistically match similar observations in the treatment and control groups based on their characteristics. In addition, the comparison of outcomes for client groups receiving different levels of program services helped to clarify the distinction between program impacts and differences related to selection bias. Despite these measures, the results are likely to contain some degree of selection bias. This weakness is discussed in more detail in the next section.

There were several other advantages of the research approach. The mixed-method approach led to an expansion in both quantitative and qualitative knowledge about entrepreneurs and their microenterprises. The studies provided comprehensive portraits of entrepreneurial households in three countries and described their management of enterprises within household economies. The implementation of the same research approach in India, Peru, and Zimbabwe provided a rare opportunity to compare findings based on the application of similar methods in very different settings. Because the studies included extensive contextual analysis, it was possible to formulate some preliminary conclusions about how place and program affect impacts. These types of conclusions are useful to both donors and program managers and can be used to improve resource allocation and program design.

B. Disadvantages of the Research Strategy and Implications for Future Research

It is important to note that the research approach has several disadvantages. These methodological limitations should be kept in mind when interpreting the results of the AIMS core impact assessments and when planning future studies. Each of these limitations represents a compromise involving practical constraints and an assessment of the tradeoffs between alternative methodologies.

1. Elimination of Selection Bias

Perhaps the most important limitation of the research strategy is that it does not eliminate all possible selection bias, leading to results that may overestimate the positive impacts of microenterprise services and underestimate the negative impacts. The ideal approach for eliminating selection bias in social science research is to use an experimental design. However, the random selection of qualified applicants to receive or not receive program services raises ethical and public relations issues and is generally rejected by program managers. Because of these objections, an experimental design was not possible for the AIMS studies. Instead, the

research followed a quasi-experimental design, along with the use of panel data and an ANCOVA estimation procedure.

While the objective of an impact assessment is to measure the impact of the program on the outcome variables, there are four other types of variables that can affect outcomes: 1) observed time-variant variables, 2) observed time-constant variables, 3) unobserved time-variant variables, and 4) unobserved time constant variables. The ANCOVA approach implemented here controls for the effects of the observed variables by explicitly including them as covariates in the estimation procedure. The unobserved variables are not directly included in the analysis, but the time-constant unobserved variables are indirectly included in the analysis through their influence on the baseline levels of the outcome variables.

An alternative to the procedure used in the CIA would be to estimate a fixed-effects model, either in conjunction with ANCOVA or as a simple fixed-effects model. The fixed effects approach could be used to “sweep out” both observed and unobserved time-invariant variables, but it would not control for unobserved, time-variant variables.²³ Because the fixed effects model is based solely on differences (changes) between the two rounds of the survey, one disadvantage of this approach is that it would not utilize information on different starting levels for the outcome variables or different levels of participation in the program prior to the baseline. The observed time-invariant variables are differenced out in the fixed effects approach, which leads to results that do not provide empirical information on the relationships between these variables, such as gender and enterprise sector, and the outcome variables.

2. Use of Pre-Treatment Measures in Baseline

A second important limitation of the methods used in the core impact assessments is that the baselines are not true pre-treatment measures of the outcome variables. In other words, the baseline measures were taken after the client groups had received some program services. Therefore, some of the positive and negative impacts of microenterprise services may already be present in the baseline measures. To the extent that some impacts are already reflected in the baseline measures, then the impact results may underestimate both positive and negative impacts.

In order to conduct an impact study with a true pre-treatment baseline, it is necessary to collect data on incoming clients in the baseline. The logistical challenge with this approach would be to collect the baseline data between the moment when clients are approved for the program and the moment when they receive their first program services. For many microfinance organizations, this window of opportunity lasts for only a few days. If the client sample consisted exclusively of incoming clients, then a longitudinal study of only two or three years would not provide the information needed to reach conclusions about long-term impacts.

Note that the lack of a pre-treatment measure may remove some of the effects of selection bias on positive impacts since these two influences may work in opposite directions. In other words, the lack of a true baseline may lead to underestimation of positive impacts while selection bias may lead to overestimation of positive impacts. On the other hand, both limitations can

²³ The problem of accounting for the unobserved time-variant variables can be partially addressed with additional assumptions and a two-step estimation procedure.

contribute to underestimation of negative program impacts. Consequently, the approach used in this research may seriously underestimate the magnitude of any negative impacts, and an important advantage of having a pre-treatment baseline would be that the study would be more likely to uncover the existence of negative impacts.

The interaction between selection bias and lack of a pre-treatment baseline not only creates the potential for opposite biases in the impact results, but it also motivates substantial interest in analyzing impacts on new entrants to the program. The new entrants are the people who were originally in the control group but who received program services for the first time between the baseline and second round of the survey. There are several advantages in comparing the new entrants to the control group. First, their baseline measures on the outcome variables, taken before they received program services, represent true pre-treatment measures. If their baseline measures are similar to those of the control group, then that provides some evidence against the presence of substantial selection bias.²⁴

3. Measurement Issues

A third important disadvantage of the research approach was that it relied on a relatively unsophisticated measure of the level of program participation. Due to limitations on the type and quality of client credit data available, it was necessary to use a weak definition of the treatment dosage. For the treatment group, the dosage was defined as having received program services at least two years before the second-round survey, even though the length and depth of prior participation may have varied dramatically across members of the treatment group. This particular treatment dosage was selected because the length of time (two or more years) was considered sufficient for many changes to be measurable and because program participants could be reliably identified at the time of the baseline. It was difficult to obtain complete and consistent data on other, more detailed measures of program participation.

The three studies partially compensated for the weakness in the definition of the treatment dosage, but each in different ways. In the Zimbabwe studies, separate subgroups of clients were identified based on the number of loans received prior to the baseline and whether they continued to receive loans after the baseline interview. The India study included an analysis of impacts related to the cumulative number of loans taken. In addition, the India study included a unique subgroup of SEWA Bank clients who were savers but did not have outstanding loans at the time of the baseline survey. In Peru, an alternative analysis was conducted using length of time in the program as the dosage variable. In addition, a separate impact analysis was conducted on those who had received their first loans less than two years before the 1999 survey. The results of these different analyses provided some additional insights into the relationship between changes in the outcome variable and level of program participation.

Future impact studies should attempt to use more sophisticated measures of program participation. This would expand the range of possible analytical approaches and, hopefully, provide useful insights into the relationships between different levels of program participation and the impacts associated with them. Unfortunately, detailed information on program participation will continue to be limited to what microfinance organizations routinely collect as

²⁴ It is possible, however, that later entrants to a program are dissimilar from the early joiners.

part of their management information systems. At this time, few organizations have the ability to retrieve complete transaction histories for each individual client. If that information were available, then better measures of program participation would be possible.

Other limitations of the research approach related to the weaknesses of some of the impact indicators used. In particular, some of the indicators for measuring attitudinal and psychological variables were not sensitive enough to detect the occurrence of subtle changes. As a result, it was not possible to adequately test many of the individual-level hypotheses with the survey data, and it became necessary to rely more heavily on the qualitative findings. Additional problems with data quality related to empirically defining the impact variables for measurement within the local contexts. This was the case with measuring assets in India and measuring income in Zimbabwe. In general, however, the data quality in all three studies was high.

C. Conclusion

In closing, the results of the AIMS core impact assessments contribute significantly to the available information about the client-level impacts of microenterprise support programs. The research strategy used in these three impact assessments has both advantages and disadvantages. It represents a methodological advance in the development of rigorous impact assessments within this field. As indicated in the previous section, the research strategy reflects practical considerations and tradeoffs that researchers designing future impact studies may want to reassess. In particular, the implementation of an experimental design with a true pre-treatment baseline and a sophisticated measure of program participation would push the frontier of knowledge out considerably.

Finally, there is a need for impact assessments that look beyond the household to market-level, regional, and macroeconomic impacts. While beyond the scope of the AIMS core impact assessments, the measurement of impacts at these levels would complete the impact picture. Household-level impact studies may understate the negative impacts of microfinance by failing to account for possible market displacement of non-clients. On the other hand, possible positive impacts on the regional economy due to multiplier effects of income and employment are also disregarded in research at the household level. By combining market-level and regional impact information with the detailed household- and enterprise-level results of studies like the AIMS core impact assessments, it would be possible to develop a comprehensive understanding of the costs and benefits of microenterprise programs. The accumulation of information on the positive and negative impacts of microfinance programs in different contexts and settings could be used to improve the efficacy of these programs in achieving economic development objectives.

REFERENCE LIST

- AIMS Team. 2001. *Conceptual Framework for Assessing the Impacts of Microenterprise Services*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- . 1997. Research Plan for the AIMS Core Impact Assessments. Unpublished AIMS Project Paper, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- . 1996. Core Impact Assessment Strategy for the Microenterprise Impact Project. Unpublished AIMS Project Paper, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Barnes, Carolyn. 2001. *Microfinance Program Clients and Impact: An Assessment of Zambuko Trust, Zimbabwe*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Barnes, Carolyn and Erica Keogh. 1999. *An Assessment of the Impact of Zambuko's Microenterprise Program in Zimbabwe: Baseline Findings*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Benus, Jacob M., Michelle L. Wood and Neelima Grover. 1994. *Self-Employment as a Reemployment Option: Demonstration Results and National Legislation*. Unemployment Insurance Occasional Paper 94-3. Washington, D.C.: U.S. Department of Labor, Employment and Training Administration.
- Burgess, R. D. 1989. Major Issues and Implications of Tracing Survey Respondents. Chapter in *Panel Surveys*, edited by D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh. New York: Wiley.
- Chen, Martha Alter. 1997. *A Guide for Assessing the Impact of Microenterprise Services at the Individual Level*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Chen, Martha Alter and Elizabeth Dunn. 1996. *Household Economic Portfolios*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Chen, Martha Alter and Donald Snodgrass. 2001. *Managing Resources, Activities, and Risk in Urban India: The Impact of SEWA Bank*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- . 1999. *An Assessment of the Impact of SEWA Bank in India: Baseline Findings*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.

- Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally College Publishing Company.
- Davis, James A. 1985. *The Logic of Causal Order*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-055. Beverly Hills: Sage Publications.
- Dunn, Elizabeth. 1999. *Microfinance Clients in Lima, Peru: Baseline Report for AIMS Core Impact Assessment*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- . 1997. *Research Plan for the AIMS Core Impact Assessments*. Unpublished AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Dunn, Elizabeth and J. Gordon Arbuckle Jr. 2001 *The Impacts of Microcredit: A Case Study from Peru*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Frison, L. and S. J. Pocock. 1992. "Repeated Measures in Clinical Trials: Analysis Using Mean Summary Statistics and its Implications for Design." *Statistics in Medicine* 11: 1685-1704.
- General Accounting Office. 1990. *Case Study Evaluations*. Washington, D.C.: United States General Accounting Office, Program Evaluation and Methodology Division.
- Godsey, Larry D. 1996. *Selecting Indicators of Sustainable Farming Systems*. Unpublished M.S. thesis. Columbia, MO: University of Missouri, Department of Agricultural Economics.
- Hill, M. O. 1973. "Diversity and Evenness: A Unifying Notation and its Consequences." *Ecology* 54(2): 427-432.
- Hulme, David and Paul Mosley. 1996. *Finance Against Poverty*. New York: Routledge.
- Khandker, Shahidur R. 1998. *Fighting Poverty with Microcredit: Experience in Bangladesh*. New York: Oxford University Press.
- Lapar, Ma. Lucila A., Douglas H. Graham, Richard L. Meyer and David S. Kraybill. 1995. "Selectivity Bias in Estimating the Effect of Credit on Output: The Case of Rural Nonfarm Enterprises in the Philippines." Economics and Sociology Occasional Paper No. 2231. Columbus, OH: Ohio State University, Department of Agricultural Economics.

- Liker, Jeffrey, Sue Augustyniak, and Greg J. Duncan. 1985. "Panel Data and Models of Change: A Comparison of First Difference and Conventional Two-Wave Models." *Social Science Research* 14: 80-101.
- Markus, Gregory B. 1979. *Analyzing Panel Data*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-018. Beverly Hills: Sage Publications.
- Menard, Scott. 1991. *Longitudinal Research*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-076. Beverly Hills: Sage Publications.
- Moffitt, Robert. 1991. "Program Evaluation with Non-Experimental Data." *Evaluation Review* 15(3): 291-314.
- Morduch, Jonathan. 1999. Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh. Unpublished manuscript. Cambridge, MA: Harvard University, Department of Economics.
- Pitt, Mark M. 1999. Reply to Jonathan Morduch's "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh." Unpublished manuscript. Providence, RI: Brown University, Department of Economics.
- Pitt, Mark M. and Shahidur R. Khandker. 1998. The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter? *Journal of Political Economy* 106(5).
- . 1996. *Household and Intrahousehold Impact of the Grameen Bank and Similar Targeted Credit Programs in Bangladesh*. World Bank Discussion Paper 320. Washington, D.C.: World Bank.
- Reichardt, Charles S. 1979. The Statistical Analysis of Data from Nonequivalent Group Designs. Chapter in *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, edited by Thomas D. Cook and Donald T. Campbell. Chicago: Rand McNally College Publishing Company.
- Reichardt, Charles S. and Harry F. Gollob. 1997. When Confidence Intervals Should be Used Instead of Statistical Significance Tests, and Vice Versa. Chapter in *What If There Were No Significance Tests?*, edited by Lisa L. Harlow, Stanley A. Mulaik and James H. Steiger. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Reichardt, Charles S. and Melvin M. Mark. 1998. Quasi-experimentation. Chapter in *Handbook of Applied Social Research Methods*, edited by Leonard Bickman and Debra J. Rog. Thousand Oaks, CA: Sage Publications.
- Rossi, Peter H. and Howard E. Freeman. 1989. *Evaluation: A Systematic Approach*. Newbury Park, CA: Sage Publications.

- Sebstad, Jennefer, and Monique Cohen. 2000. *Microfinance, Risk Management, and Poverty*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Sebstad, Jennefer and Gregory Chen. 1996. *Overview of Studies on the Impact of Microenterprise Credit*. AIMS Project Report, USAID/G/EG/MD. Washington, D.C.: Management Systems International.
- Sebstad, Jennefer, Catherine Neill, Carolyn Barnes, and Gregory Chen. 1995. *Assessing the Impacts of Microenterprise Interventions: A Framework for Analysis*. USAID Managing for Results, Working Paper No. 7. Washington, D.C.: USAID.
- Singleton, Royce A., Bruce C. Straits, and Margaret Miller Straits. 1993. *Approaches to Social Research*. New York: Oxford University Press.
- World Bank. 2000. *World Development Report 2000/2001: Attacking Poverty*. Washington, D.C.: World bank
- Yin, Robert K. 1994. *Case Study Research: Design and Methods*. Second edition. Thousand Oaks, CA: Sage Publications.
- Zeller, Manfred, Akhter Ahmed, Suresh Babu, Sumiter Broca, Aliou Diagne and Manohar Sharma. 1996. *Rural Financial Policies for Food Security of the Poor: Methodologies for a Multicountry Research Project*. Food Consumption and Nutrition Division Discussion Paper No. 11. Washington, D.C.: International Food Policy Research Institute.

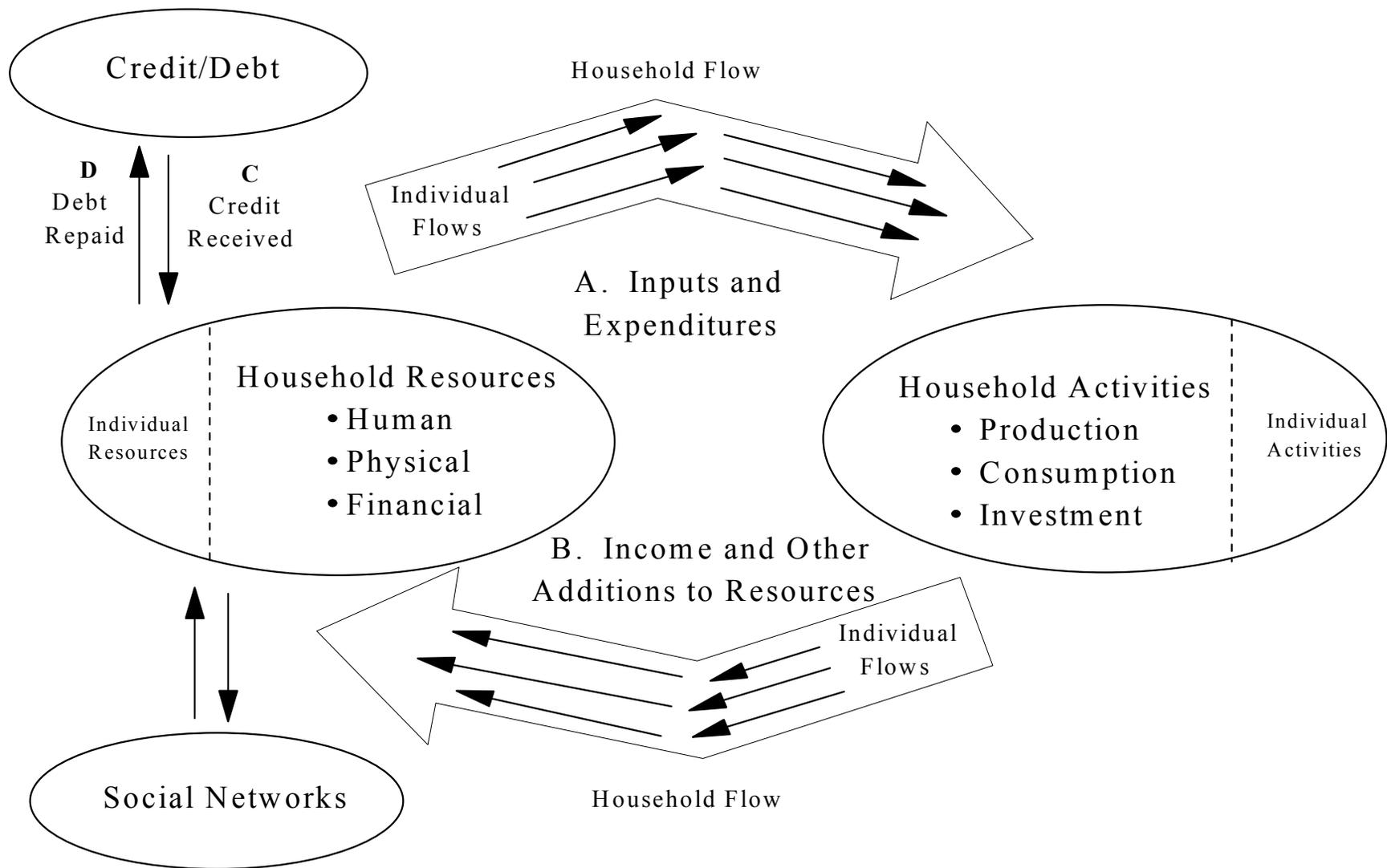


Figure 1: Conceptual Model of the Household Economic Portfolio

Table 1: Household-Level Impact Variables

Hypothesis and Variable	Measure	Scale	Moderating Variables
H-1. Household Income	Income received by all HH members in previous year (Z: previous month) from all sources (for ME income, includes net income or profit)	ratio	number of economically active HH members ^c ; number of HH income sources
H-2. Diversification of Income	Inverse Simpson index ^a	interval	number of economically active HH members
H-3a. Housing Improvements	Expenditures on building materials (used and unused) and labor payments in previous 12 months for housing improvements, repairs, expansions, and infrastructure connections (Z: does not apply to lodgers)	ratio	housing tenure
H-3b. Appliances and Transport	Expenditures on HH appliances and vehicles in previous 24 months (I/Z: also includes furniture)	ratio	household life-cycle (age/average age of head/heads of household)
H-3c. ME Fixed Assets	Current aggregate monetary value of fixed assets in ALL MEs associated with the HH	ratio	number of MEs
H-4. Children's Education	P: Total education-related expenditures per student in current calendar year; I/Z: Percentage of children in relevant age range currently enrolled in school	ratio	
H-5. Food Expenditures	I/P: Daily per capita expenditures on food and beverages consumed in and out of the home; Z: Number of times certain food items (meat, eggs, fruit) eaten in previous week	ratio	
H-6. Coping with Shocks	Type of coping mechanisms used in dealing with most damaging shock in previous 2 years (0=used at least some stage II strategies; 1=used only stage I strategies) ^b	nominal	

Abbreviations: P=Peru; I=India; Z=Zimbabwe; ME=microenterprise; HH=household

Notes: ^a The inverse Simpson index is calculated as $1/q$, where $q = p_1^2 + p_2^2 + p_3^2 + \dots + p_N^2$. Each p_i is the proportion of household income generated by the i^{th} source of income and N is the number of income sources. The index equals 1 when there is one source of income, and it equals N when there are N sources that each contribute equal amounts ($1/N$) to income (Hill 1973).

^b A stage II strategy is defined as the loss of use of a productive (income-generating) asset.

^c Included household members who earn a wage/salary or manage/work in the household's enterprises (whether paid or unpaid).

Table 2: Enterprise-Level Impact Variables

Hypothesis and Variable	Measure	Scale	Moderating Variables
E-1. Microenterprise Revenue	Gross sales revenue in previous month (in addition: Z: net revenue in previous month; P: net revenue in previous year)	ratio	sector; location ^b ; gender of entrepreneur
E-2. Enterprise Fixed Assets	Current monetary value of all fixed assets used in ME	ratio	sector; location; gender of entrepreneur
E-3. Employment	1) Total number of hours worked by proprietor and paid and unpaid employees in previous week; 2) Total (aggregate) number of days worked by proprietor and paid and unpaid employees in previous month; 3) (optional) Monetary value of wages paid (including in-kind payments) in previous week and month	ratio	sector; location; gender of entrepreneur
E-4. Transactional Relationships	1) Main type of suppliers (0=individuals or retailers; 1=wholesalers, intermediaries, or manufacturers); 2) Main type of customers (0=final consumers; 1=retailers, wholesalers, intermediaries, or manufacturers); (Z: not available) 3) (optional) Marketing margin ^a (for commercial sector); 4) (optional) Tenure of business premise (0=insecure; 1=secure); 5) (optional) Type of premise (context specific measures); 6) (optional) Customer credit (0=does not extend customer credit; 1=does extend customer credit); 7) (optional) Fixed sales contracts (0=does not have fixed sales contracts; 1=does have fixed sales contracts)	nominal nominal ratio nominal nominal nominal nominal	sector; location; gender of entrepreneur

Abbreviations: P=Peru; I=India; Z=Zimbabwe; ME=microenterprise; HH=household

Notes: ^a The marketing margin, or percentage mark-up, is calculated as (sale price - purchase price)/(purchase price).

^b More than one location-related moderating variable may be relevant, depending on the conditions at the research site. For example, location may refer to geographic location, when the sample covers distinct areas, or it may refer to commercial location (i.e., home-based, market-based, commercial storefront).

Table 3: Individual-Level Impact Variables

Hypothesis and Variable	Measure	Scale	Moderating Variables
I-1. Control over Resources and Income within HHEP	Three decision variables: 1) (for clients only) Who decided to take last loan? ^a 2) (for clients only) Who decided how to spend last loan? 3) Who decided how to spend ME revenue? Response categories: ^b 0=others make the decision; 1=respondent makes the decision.	nominal	gender; marital status; presence of additional income earners in HH
I-2. Self-Esteem and Respect	Two variables: 1) Do you feel you feel you make an important contribution to the HH? 2) Do you feel that the adult members of your HH respect the contributions you make to the HH? Response categories: 0=negative responses; 1=positive responses.	nominal	gender; marital status
I-3. Personal Savings	I/P: Do you have personal savings? Z: Do you have a personal savings account? Responses: 0=no; 1=yes.	nominal	gender; marital status; household income level ^c
I-4. Position to Deal with Future	1) Do you feel you are prepared, or in a good position to deal with the future? Responses: 0=negative responses; 1=positive responses. 2) Are you doing anything to prepare yourself for the future? Responses: 0=no; 1=yes.	nominal	gender; household income level; presence of additional income earners in HH

Abbreviations: P=Peru; I=India; Z=Zimbabwe; ME=microenterprise; HH=household

Notes: ^a The two questions related to the last loan should only be analyzed if a loan was received subsequent to the one discussed in the baseline survey.

^b Any decision in which the respondent participates (including joint decisions) should be coded as “1”.

^c Household income as a moderating variable is meant to reflect socioeconomic level. If an actual income measure is unavailable (e.g., too many missing values), then an appropriate numeric or categorical proxy for household income can be used.

Table 4. Study Questions and Research Propositions for Case Studies

Focus	Study Question	Research Proposition
Components of the Household Economic Portfolio	Q1: Have the components of the household economic portfolio changed? If so, how, and what factors led to these changes?	P1: Participation in program services relieves a capital constraint on the household economic portfolio, which allows the household to increase one or more of its production, consumption, or investment activities. Some households will use loans to increase the base of income generating activities, leading to an increase in the flow of income from activities to resources and a net increase in the financial resources available to the household.
Financial and Risk Management Behavior	Q2: Has the financial and risk management behavior of the household or its members changed? If so, how, and what led to these changes?	P2: Participation in program services improves the financial management options available to the household by offering a reliable source of borrowed funds. Participation in program services allows households to accumulate and maintain savings and other near-liquid forms of assets, increasing the effectiveness of stage I coping strategies and helping to avoid stage II coping strategies.
Intrahousehold Control Over Resources	Q3: Have the patterns for intrahousehold control over resources and income changed? If so, how, and what factors led to these changes?	P3: Female partners are more likely to have control over the decisions related to applying for and spending a program loan if they are the named clients. Loans allocated to an income generating activity managed by the female partner will help her increase income from her activity, and her influence over the allocation of other household resources will increase over time.
Changes in Self Esteem	Q4: Have changes occurred in the client's self-esteem? If so, how, and what factors led to these changes?	P4: The receipt and use of the credit allows the client to increase his or her contribution to the material welfare of the household and the community and become better managers of resources. The client's self-esteem increases with these positive changes.
Changes in Orientation Toward the Future	Q5: Have changes occurred in the client's orientation toward the future? If so, how, and what led to the changes?	P5: The availability of a steady and reliable source of credit leads the client to have a more positive orientation toward the future in the sense that the client is better able to formulate and more effectively implement proactive financial and economic plans.
Changes in Transaction Relationships	Q6: Have changes occurred in the transaction relationships of the microenterprise? If so, how, and what factors led to these changes?	P6: Participation in program services relieves a capital constraint on the enterprise, which improves the ability of the enterprise to buy inputs in bulk, reach new input and output markets, and maintain a higher and/or more reliable flow of outputs.
Reported Impacts and Client Perspective	Q7: What changes does the client perceive as a result of program participation, and how did those changes occur? What value does the client place on the changes, program services?	P7: Clients perceive positive economic and social changes in their lives as a result of program participation. Clients value loans as useful tools that help them to reach economic goals, thereby increasing the economic well being of their households.

APPENDIX: Calculating Gain Scores

Suppose that there are N_P participant households in the study and N_C control households. Let X_{it} denote the measurement on the outcome variable (X) for a participant household, where i represents a particular participant household ($i=1, \dots, N_P$) and t represents a particular round of the survey ($t=1,2$). Similarly, let X_{jt} denote the measurement on the outcome variable for a control (non-participant) household (where $j=1, \dots, N_C$). The gain score analysis of impact is based on the comparison of O_P , the average change in the outcome variable for the participant households, to O_C , the average change in the outcome variable for the control households, where

$$\bar{X}_P = \frac{\sum_{i=1}^{N_P} (X_{i2} - X_{i1})}{N_P} \quad \text{and} \quad \bar{X}_C = \frac{\sum_{j=1}^{N_C} (X_{j2} - X_{j1})}{N_C}$$

Operationally, this involves three steps: 1) calculate the change in the outcome variable for each household; 2) find the mean values of the change scores for the participant subgroup and the control subgroup; and 3) compare the mean values and test for statistically significant differences between them. The first step, calculation of the change scores for each household, involves subtracting the measurement on the outcome variable in the first round of the survey from the measurement on the outcome variable in the second round ($X_{*2} - X_{*1}$). This operation is repeated for each household so that there is a raw change score associated with each household.²⁵ The mean values of change can be calculated using the above formulae for O_P and O_C , after separating the participant data from the control group data.

The same operational procedures can be used to calculate change scores for the nominal variables. Note that all of the impact variables measured on the nominal scale are dichotomous, with a value of zero assigned to a “bad” outcome and a value of one assigned to a “good” outcome. For these nominal variables, there are three possible outcomes for the individual change scores:

- 1) If response went from a bad outcome in the baseline survey ($X_{*1} = 0$) to a good outcome in the second-round survey ($X_{*2} = 1$), then the change score would be $X_{*2} - X_{*1} = 1 - 0 = 1$. A measure of “1” for the change score represents a positive impact.
- 2) If response went from a good outcome in the baseline survey ($X_{*1} = 1$) to a bad outcome in the second-round survey ($X_{*2} = 0$), then the change score would be $X_{*2} - X_{*1} = 0 - 1 = -1$. A measure of “-1” for the change score represents a negative impact.
- 3) If response did not change between the two rounds of the survey, then the change score would be 0, which represents no impact. The two possible cases would be a) $X_{*1} = X_{*2} =$

²⁵ For simplicity, this discussion focuses on the household-level impact variables. For the enterprise-level variables, there would be a raw change score associated with each enterprise in the data set. Likewise, raw change scores would be associated with individual respondents for the individual-level variables.

1, in which case the change score would be $1 - 1 = 0$, or b) $X_{*1} = X_{*2} = 0$, so that the change score would be $0 - 0 = 0$.

Participant Data			Control Group Data		
X_{i2}	X_{i1}	Gain Score	X_{j2}	X_{j1}	Gain Score
1	1	0	0	0	0
1	0	1	1	0	1
0	0	0	0	1	-1
1	0	1	0	0	0
1	1	0	1	1	0

Figure 4: Example of Gain Scores for Nominal Data

An example of the computation of gain scores for nominal data is provided in figure 4. The participant data are provided in the first three columns, while the control group data are in the next three columns. The data are presented in the following order: second-round measure, baseline measure, and gain score. In this example, the mean values of change are $0_p = (0+1+0+1+0)/5 = .20$ and $0_c = (0+1-1+0+0)/5 = 0$.