



USAID
FROM THE AMERICAN PEOPLE

MANUAL FOR THE IMPLEMENTATION OF USAID POVERTY ASSESSMENT TOOLS

PREPARED BY

THE IRIS CENTER, UNIVERSITY OF MARYLAND - COLLEGE PARK

UPDATED: MAY 2011

DISCLAIMER

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

PREPARED BY

The IRIS Center, University Research Corporation International

CONTACT INFORMATION

The IRIS Center
University of Maryland
3106 Morrill Hall
College Park, MD 20742
USA

E-mail: info@iris.umd.edu
Phone: +1.301.405.3110
Fax: +1.301.405.3020
Web: www.iris.umd.edu

CONTENTS

CHAPTER ONE – INTRODUCTION	7
Purpose of this Manual	7
What is a PAT?	7
Why PAT? - U.S. Legislation and USAID Requirements	8
Tool Development	9
Poverty Lines – what does ‘very poor’ mean?	9
Overview of Implementation Process	10
How to Use This Manual	10
Note on Survey Methodology	11
Case Study: The Mingunesian Microfinance Association	12
CHAPTER TWO – PLANNING AND PREPARATION FOR IMPLEMENTATION	13
Scheduling	13
Budgeting	14
Personnel	15
Logistics	18
Equipment	18
Workload	19
CHAPTER THREE – SAMPLING	20
Target population – Who do you want to learn about?	21
Sampling methodology – Simple Random Sampling (SRS)	22
Sampling methodology – Cluster Sampling	24
SRS vs. Cluster Sampling – an example in pictures	30
Stratified Sampling	31
Sample Selection	32
Using a random number table	32
Interval selection at the cluster level (Systematic Probability Proportion to Size (PPS) Sampling)	34

Interval selection at the client/beneficiary level (Systematic Random Sampling)	36
CHAPTER FOUR – PLANNING FOR FIELD WORK	37
Preparing the PAT Survey	37
Translation	37
Printing & Copying	38
Developing a Survey Tracking System	39
Scheduling Interviews	39
Field Preparation	41
Training Interviewers	42
Before the Interview	43
Conducting the Interview	43
Ending the Interview	44
Do’s and Don’ts	44
Sample Interviewer Training Schedule	45
Practice Interviews / Pre-testing the Survey	45
CHAPTER FIVE – FIELD WORK: INTERVIEWING AND QUALITY CONTROL	46
Field Supervision	46
Time Management	46
Quality Control	47
Document Control and Tracking	48
Internal Audit	48
CHAPTER SIX – DATA PROCESSING - EPI INFO AND CSPO	50
Quality Control at Main Office	50
Epi Info	51
What is Epi Info?	51
System Requirements	52
Saving Files	52
First Steps	52
Basic Functions	53
Make View	54
Enter Data	56
Analyze Data	66
Merging Data Sets	67
Frequencies (looking for odd / outlying values)	70
Finding specific survey data	72

Calculating the Poverty Level	74
Troubleshooting	78
CSPro	80
What is CSPro?	80
System Requirements	80
Installation	80
Data Entry Application Overview	82
Accessing Data That Has Been Saved	86
Modifying Data and Adding New Cases	88
Compare Data Tool	89
Tabulate Frequencies Tool (Looking For Odd Values)	93
Merging Data (Concatenate Data Tool)	96
Tabulating Calculation of the Poverty Level	98
Export Data Tool	100
Reporting to USAID	102
CHAPTER SEVEN – ADDITIONAL RESOURCES	103
PAT Website: www.povertytools.org	103
Online Training	103
PAT Help Desk: pathelp@iris.umd.edu	103
Regional Trainings	103
APPENDIX A – IMPLEMENTATION TEAM JOB DESCRIPTIONS	105
Project Manager	105
Field Supervisors	106
Interviewers	107
Sampling and Survey Tracking Coordinator	108
Data Processing Coordinator	109
Data Processors	110
APPENDIX B – GUIDE FOR INTERVIEWER DEBRIEFING DISCUSSIONS	111

CHAPTER ONE – INTRODUCTION

Purpose of this Manual

This manual was written to provide technical guidance on how to implement a USAID Poverty Assessment Tool (PAT). Specifically, it was written to help microenterprise practitioners meet the reporting requirements of the 2000 Microenterprise for Self-Reliance and International Anti-Corruption Act.

The following pages detail every step of the PAT implementation process, from understanding the purpose and design of the PATs, through developing an implementation plan, developing a sampling plan, choosing and training an implementation team, conducting interviews, and entering and analyzing data. While some guidance given will apply specifically to required USAID implementers, the majority of the information will be useful to anyone interested in using a PAT.

See page 10 for information on how to use this manual, including guidance on the sections that will be most beneficial to particular members of your PAT implementation team.

What is a PAT?

Each USAID Poverty Assessment Tool (PAT) consists of a short, country-specific Survey and a Data Entry Template. PATs were designed to measure the prevalence of poverty and extreme poverty of households in a group. Each PAT survey is:

- Country-specific
- 10-25 questions in length
- Based on nationally representative household data
- Used to estimate a program's outreach to poor and very poor households
- Easy to use¹

As with any poverty tool, the PAT cannot predict with a strong degree of accuracy whether one particular household is above or below a poverty line. Thus, the current PAT toolkit is not intended to be used as a tool for targeting households at a certain poverty level. Instead, the PATs were designed to maximize accuracy at the aggregate level, by balancing errors resulting in misclassification of households. Also, while the PAT can be used to measure the prevalence of poverty of a group at different points in time, this should not be confused with assessing impact on poverty of a particular program. A proper impact assessment requires a more involved process to determine attribution.

PAT surveys are administered according to standard sampling and interviewing protocols, detailed in this manual, and the resulting data are entered and analyzed using the corresponding Data Entry Template. Each Data Entry Template:

¹ Care is given during tool development to balance accuracy of results with practicality of administering the questions. This will be discussed in further detail starting on page 9.

- Is country-specific
- Works within either the Epi Info or CSPro Software Package (see page 50)
- Includes a program to automatically calculate the poverty prevalence of the target population using one or more pre-determined poverty lines
- Is customizable (for translations, running additional analyses, etc.)

Why PAT? - U.S. Legislation and USAID Requirements

Microenterprise development programs have been among the most promising donor-sponsored programs for improving the lives of poor people. In 2000, the U.S. Congress passed the Microenterprise for Self-Reliance and International Anti-Corruption Act, which mandated that **half** of all USAID microenterprise funds benefit the very poor. To verify that USAID meets this target, subsequent legislation requires USAID to develop and certify low-cost tools for assessing the poverty status of microenterprise beneficiaries², and to require its microenterprise implementing partners to use those tools to measure and report the share of their beneficiaries who are very poor. USAID PATs were the resulting tools designated to fulfill this requirement, allowing implementing partners to gain an accurate estimate of the share of their beneficiaries who are very poor.³

Based on the law and on subsequent USAID policy decisions, USAID does:

- Require that organizations spending \$100,000 or more in USAID funds for microfinance or microenterprise development in a country for which a PAT is available, use that PAT at least annually to measure and report the share of their clients who are “very poor” as defined in the law; and
- Give the USAID Mission or other operating unit that funds the activity in question the authority to make “reasonable exceptions” to the general requirement to use the PAT.⁴

The law does not:

- Require any of the organizations that are partnering with USAID on microenterprise development at any given time to exclusively target 'very poor' families.
- Mandate that 50 percent of each organization's clients be among the very poor, but that 50 percent of USAID's total microenterprise development resources benefit that group.
- Mandate that enterprise development programs or policy reform activities cease or even that they necessarily be scaled back.
- Require that all microenterprise partners measure poverty levels of all clients, or even all incoming clients. A representative sample is sufficient.

² Beneficiaries include the employees of clients, so Enterprise Development/BDS programs should include the employees of their clients in their sampling plan.

³ The legislation defines 'very poor' as those earning less than US\$1.25 per day per capita -- adjusted for purchasing power parity (PPP) -- or those who are in the bottom half of those under the poverty line in their country.

⁴ The basis for such exceptions is covered in Automated Directives System (ADS) Chapter 219.

Tool Development

USAID commissioned The IRIS Center at the University of Maryland to develop, test, and disseminate poverty assessment tools that meet Congressional requirements for accuracy and practicality. After extensive desk research and soliciting input from the microenterprise community, accuracy tests of poverty indicators were implemented by IRIS in Bangladesh, Peru, Uganda, and Kazakhstan in 2004. Datasets from the World Bank's Living Standards Measurement Survey were also analyzed to identify the most accurate poverty indicators in eight additional countries (Albania, Ghana, Guatemala, India (Bihar and Uttar Pradesh only), Jamaica, Madagascar, Tajikistan, and Vietnam).

From this initial round of research, IRIS and USAID determined that international or regional tools would not be sufficiently accurate, and that at least one tool would need to be certified for each of the countries that receive USAID microenterprise funding (and that tools will need to be recalibrated periodically). USAID and IRIS worked with practitioners to develop certification criteria for the country-specific tools, incorporating both accuracy and practicality.

The field tests for practicality were carried out in late 2005 and early 2006. Combining the accuracy and practicality results, IRIS developed a country-specific tool for each of the 12 countries involved in the research, and continues to develop new country-specific tools, with 29 PATs available as of May 2010. (See www.povertytools.org for an updated list of PAT availability.)

Poverty Lines – what does ‘very poor’ mean?

As stipulated in the previously mentioned U.S. legislation, PATs are designed to assess the poverty level of groups with respect to either the national or the international extreme poverty line.

The national extreme poverty line (also referred to as the ‘national median poverty line’) divides those below the national poverty line for a given country in half. Those in the bottom half are considered ‘very poor’ or ‘extremely poor’.⁵ The national extreme poverty line is NOT half of the value of the national poverty line. For example, if there are 4,000 households below the national poverty line in a country, the poorest 2,000 are considered ‘very poor’.

The international extreme poverty line – commonly called the dollar a day line – is currently equal to US\$1.25, adjusted for Purchasing Power Parity (PPP).

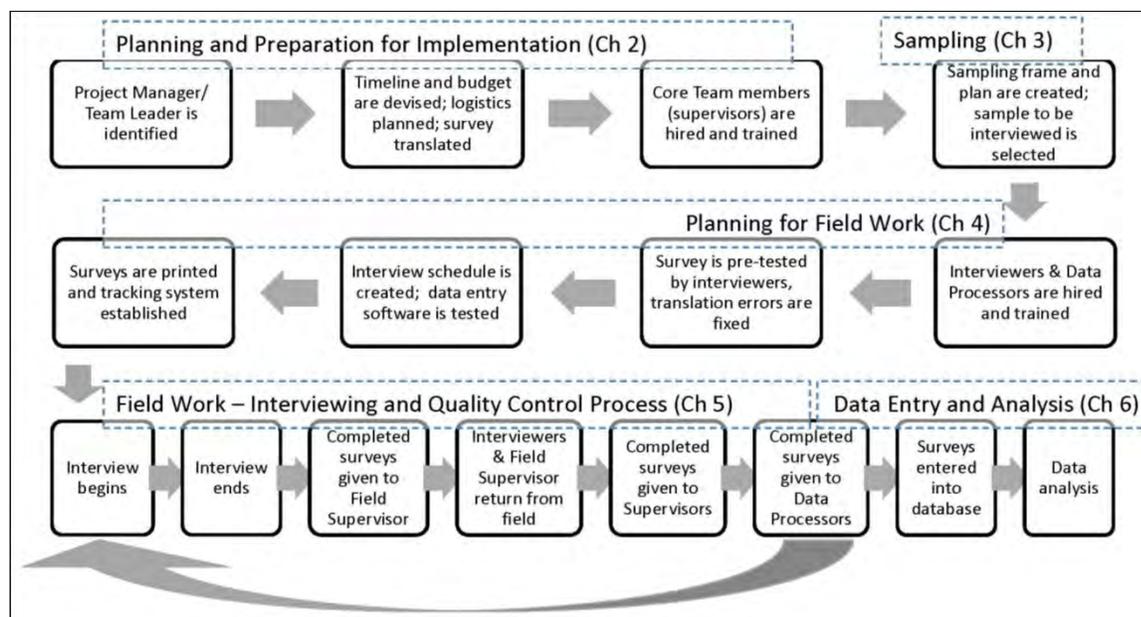
Tool developers choose the higher of these two lines – the one that is more inclusive – to include in a given country's PAT. When the poverty calculation is run, the percentage of very poor clients is calculated, using the chosen poverty line as described.

⁵ We will use these terms interchangeably: ‘very poor’ and ‘extremely poor.’

As of April 2010, additional poverty lines are being added to each PAT to provide additional means for analysis. The national poverty line is being added to those tools that use the national median poverty line for reporting. A second poverty line (US\$2.50 adjusted for PPP) is being added to tools that use the international extreme poverty line for reporting.

Overview of Implementation Process

The diagram below outlines the PAT implementation process. The headings in the dotted outlines correspond to the chapters of this manual.



How to Use This Manual

Individuals who plan on leading a team through PAT implementation or are interested in the PATs for academic or research purposes will find it most useful to read through the entire manual. Otherwise, the following reading is recommended, based on your planned involvement in the PAT implementation. For detailed descriptions of each of the roles listed, see page 15. We will refer to these same roles (job titles) throughout the manual.

<i>If you are a...</i>	<i>Read...</i>	<i>And adapt appropriate sections for use in training...</i>
Project Manager	The entire manual	Support staff (translators, logistics assistants, etc)
Field Supervisor	Chapters 1, 2, 4, 5; Appendix B; review other chapters	Interviewers
Sampling and Survey Tracking Coordinator	Chapters 1, 2, 4, 5; review other chapters	Other team members, according to their role in helping you carry out the sampling plan
Data Processing Coordinator	Chapters 1, 2, 5, 6; review other chapters	Data Processors

All team members (including Interviewers) should also read the country-specific User Guides available at www.povertytools.org. These guides provide question-by-question guidance derived from user feedback and source-survey documentation as well as notes on every step of the interview process, from beginning to end.

Note on Survey Methodology

It is the recommendation of the PAT developers that whenever possible, PAT surveys be implemented as a Household Survey (HHS). Since it is conducted in the home of the client and allows for direct observation of living conditions, the HHS often will be the most accurate assessment method, and can be implemented by any organization, regardless of their intake or evaluation methods. The guidance given in this manual will assume that surveys are being administered as HHSs.

However, we recognize that because of resource or other limitations, some implementers may choose to employ other means of survey administration, such as integrating the PAT into existing monitoring and evaluation systems or conducting surveys during group meetings, at a loan office or at other locations beneficiaries are likely to visit.

To accommodate this, we will, from time to time, include footnotes and/or textboxes with special guidance for implementers utilizing these alternative methodologies. When choosing a method for conducting surveys, consider the following:

- Household surveys allow the interviewer to visually verify the answers to some survey questions. Conducting surveys outside the home precludes this.
- If your organization already has a monitoring tool that you plan to combine with the PAT, then the PAT questions must be inserted at the beginning of the monitoring tool, with the pre-existing questions following that in order to reduce bias.
- To minimize the chance that clients will bias their responses in order to qualify for a loan or other product, it is recommended that the PAT be administered only after the client receives a loan or other product or service.
- Conducting surveys outside the household can make it difficult to maintain client confidentiality. Regardless of the methodology used, it is always important to have an interview location that avoids observers and other distractions as much as possible.

Case Study: The Mingunesian Microfinance Association

Throughout the manual, we will provide examples of how an MFI from the fictitious nation of Mingonesia – the Mingunesian Microfinance Association, or MMA – would plan for its PAT implementation. Some background notes:

As a USAID-funded organization, the Mingunesian Microfinance Association is required to report the percentage of “very poor” clients it serves. To do this we will be administering a USAID Poverty Assessment Tool to a sample of clients and analyzing the results using the USAID Data Entry Template and analysis program. While this survey is being done as part of a funding requirement, it will also benefit the MMA directly by providing us with information on our clients not previously collected. This information will give us a better sense of who our clients are and how we can better serve them. Costs are listed in Mingunesian Pounds.

Founded:	1997
Programs:	Funding Farming Friends (FFF), Home Business Builders (HBB), Loans for Success (LfS)
Portfolio:	MNP 600,000 <i>[MNP = Mingunesian Pound]</i>
USAID Funding:	US\$150,980 (FY2010)
Other Funding:	Mingunesian Development Bank, MAID
Number of Loan Groups:	45
Number of Clients:	2,300
Average Loan Size:	MNP 196
BDS services:	Agriculture training

Plan Summary

Duration:	1 June – 20 July 2010 (7 weeks)
Total Cost:	MNP 15,731
Staff Involved:	Vladimir Terskin, Desta Amde, Tracy Choi, Elliott Skabler
Additional Staff Needed:	3 Field Supervisors, 8-10 Interviewers, 2-3 Data Processors, 1 Translator

CHAPTER TWO – PLANNING AND PREPARATION FOR IMPLEMENTATION

Precise planning, careful time management and diligent field supervision not only contribute greatly to holding down costs, but are also essential for carrying out an efficient and successful field survey implementation. This chapter covers several key aspects of successful field operations: Scheduling, Budgeting, Personnel, Logistics, Equipment, and Workload.

Scheduling

A time frame for starting and completing your PAT implementation, including the order in which the various activities will take place, needs to be established early in the process. The time required and the amount of overlap between activities should be carefully estimated; the implementation team should be careful not to cut corners to save time. Field operations are best scheduled to avoid major national or religious holidays, periods of bad weather (rainy seasons), or heavy workloads. Below is a list of activities for implementing the Poverty Assessment Tool, laid out in a Gantt chart for the fictitious example of the MMA (see page 12).

	June 1-7	June 8-14	June 15-21	June 22-28	June 28-5	July 6-13	July 14-20
Hire staff							
Train core staff							
Assess & order supplies							
Assess & address technology							
Budget estimates							
Translate questionnaire							
Design sample plan							
Train Interviewers							
Train Data Processors							
Make travel arrangements							
Install and test software							
Identify questionnaire storage							
Create interview schedule							
Organize surveys							
Conduct interviews in Region A							
Conduct interviews in Region B							
Conduct interviews in Region C							
Conduct interviews in Region D							
Interview quality control							
Feedback and adjustments							
Make-up or extra interviews							
Data entry							
Data cleaning / quality control							
Analyze data							

Budgeting

It is very important to review the budget regularly to ensure that cost estimates remain in line with how much is actually spent during the implementation. You should make a small contingency fund available to cover unforeseen expenses. The example below follows the fictitious example of the MMA (see page 12). The budget categories listed are representative of the actual types of costs that implementers will likely incur, but the specific costs will need to be adapted to your situation.

PROJECT STAFF							
	# of people	# of hrs/week	# of weeks	Hourly Rate (incl. benefits)	Total Cost	Covered by Organization Budget	Additional Funds Needed
Project Manager	1	17	7	MNP 22	MNP 2,618.00	MNP 1,575.00	MNP 1,043.00
Field Supervisor	3	40	3	MNP 19	MNP 6,840.00	MNP 4,752.00	MNP 2,088.00
Interviewers	8	40	3	MNP 19	MNP 18,240.00	MNP 11,400.00	MNP 6,840.00
Sampling & Survey Tracking Coordinator	1	30	6	MNP 20	MNP 3,600.00	MNP 1,980.00	MNP 1,620.00
Data Processing Coordinator	1	22	4	MNP 15	MNP 1,320.00	MNP 880.00	MNP 440.00
Data Processors	2	19	3	MNP 11	MNP 1,254.00	MNP 513.00	MNP 741.00
Logistics Assistant	2	12	7	MNP 7	MNP 1,176.00	MNP 1,176.00	MNP 0.00
				<i>Total</i>	<i>MNP 35,048.00</i>	<i>MNP 22,276.00</i>	<i>MNP 12,772.00</i>
ADMINISTRATIVE AND MANAGEMENT STAFF							
	# of people	# of hrs/week	# of weeks	Hourly Rate (incl. benefits)	Total Cost		
Branch Manager (as Translator)	4	2	1	MNP 19	MNP 152.00	MNP 152.00	MNP 0.00
Accountant	1	2	4	MNP 12	MNP 96.00	MNP 96.00	MNP 0.00
Executive Director	1	2	1	MNP 25	MNP 50.00	MNP 50.00	MNP 0.00
Receptionist/Assistant	1	2	6	MNP 7	MNP 84.00	MNP 84.00	MNP 0.00
				<i>Total</i>	<i>MNP 382.00</i>	<i>MNP 382.00</i>	<i>MNP 0.00</i>
SUPPLIES & TRAVEL							
			# of weeks	Cost per week	Total Cost		
Fuel for vehicle			3	MNP 170	MNP 510.00	MNP 0.00	MNP 510.00
Car rental			3	MNP 80	MNP 240.00	MNP 0.00	MNP 240.00
Public transportation			0.8	MNP 25	MNP 20.00	MNP 0.00	MNP 20.00
Cell phone time			7	MNP 10	MNP 70.00	MNP 70.00	MNP 0.00
Office supplies (photocopies, paper, printer cartridge, etc.)			3	MNP 200	MNP 600.00	MNP 600.00	MNP 0.00
Meals during training			2	MNP 120	MNP 240.00	MNP 0.00	MNP 240.00
Interviewer per diems			3	MNP 200	MNP 600.00	MNP 0.00	MNP 600.00
				<i>Total</i>	<i>MNP 2,280.00</i>	<i>MNP 670.00</i>	<i>MNP 1,610.00</i>
				TOTAL	MNP 37,710.00	MNP 23,328.00	MNP 14,382.00

Personnel

Skilled personnel who are well trained and motivated can strongly influence the success of your field operation. The implementation team for the PAT should include an overall Project Manager, enough Field Supervisors to provide adequate supervision of Interviewers, the number of which will depend on your sampling plan, a Sample and Survey Tracking Coordinator, a Data Processing Coordinator, and at least two Data Processors.

Outside staff can be hired temporarily to assist on the project, especially for the roles of the Interviewers and Data Processors. However, it is recommended that when possible, the Project Manager, the Field Supervisors and the Data Processing Coordinator roles be filled by in-house staff in order to facilitate an efficient and effective implementation process.

The Core Team consists of individuals with supervisory responsibilities. They are hired (or selected from current staff) and trained early in the pre-implementation process and are responsible for training other team members.

The Core Team should include the following members:

- **Project Manager:** oversees planning and implementation of the field survey and participates in training of all staff. This person should have previous experience in survey administration and staff training, should have leadership and project management experience, and should be well-versed in the local context.
- **Field Supervisors:** trains and coordinates daily activities of Interviewers, and maintains quality control in the field. They should have significant experience in conducting interviews and managing staff, as well as in-depth knowledge of local geography, culture, and languages.
- **Sampling and Survey Tracking Coordinator:** creates and maintains tracking system for survey implementation; develops and ensures sampling plan is maintained. This person should have previous experience in sample design and good overall quantitative and organizational skills. This person should be highly accessible to all staff during the interviewing phase to answer questions and adjust the sampling plan as necessary. This person may be called upon to participate in staff training.⁶
- **Data Processing Coordinator:** identifies and trains Data Processors, maintains quality control in data processing, and runs the poverty analysis program. This person should have strong IT skills and should be able to troubleshoot hardware and software problems and supervise staff. Statistical skills will be useful for conducting any additional data analysis desired by your organization. Previous experience with the Epi Info or CSPro software package used for data entry and analysis is useful but not necessary. Detailed instructions for

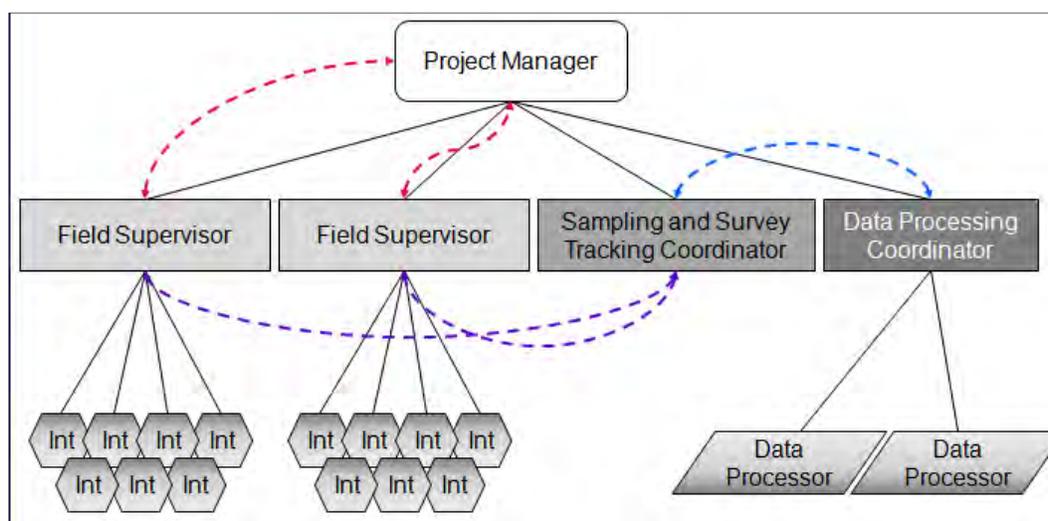
⁶ Depending on staff capacity and experience, it may be worthwhile to hire an experienced sampling consultant as the process is complex.

using the software are included in this manual and additional assistance is available via the PAT website and Help Desk.

Additional essential staff:

- **Interviewers:** thoroughly learn survey questions and interview techniques to minimize bias, and conduct interviews in a confident and relaxed manner. Interviewers with previous experience are ideal but not always available. At a minimum, they should have good interpersonal skills and fluency in the appropriate languages.
- **Data Processors:** enter and clean data collected by Interviewers. They should have strong computer skills, good attention to detail, and previous experience in data entry.
- **Translators, Logistics Assistant, additional staff** as needed⁷

The chart below shows the hierarchy of the implementation team members. The dotted lines show relationships where continuous feedback occurs though all team members should be in contact with each other throughout the implementation process. For example, the Project Manager is the supervisor of all staff but Field Supervisors provide constant feedback on the interviewing process and make recommendations to both the Project Manager and the Sampling and Survey Tracking Coordinator (SSTC) on changes to the interviewing plan and schedule. Likewise, the SSTC and Data Processing Coordinator must work together to ensure that the sampling plan is being followed and that surveys are being tracked and stored properly.



⁷ Depending on your needs and the availability of current staff, you may need to hire drivers, technical consultants, computer specialists, etc. Keep in mind that staff members who work directly with clients or beneficiaries are the ones who know their daily habits and can be quite helpful with logistical planning for field work.

Staff members must rely on each other for support and guidance throughout the survey implementation. When hiring staff, consider how well you think they will work together. The following table shows an example of how team members work together during different phases of the survey implementation.

	Team 1: Interviewers and Field Supervisors	Team 2: Sampling and Survey Tracking Coordinator, Data Processing Coordinator and Data Processors	Team 3: Project Manager and other supervisory and administrative staff
Phase One: Pre-implementation	<ul style="list-style-type: none"> • Conduct / Receive training • Test surveys • Develop interview schedule • Develop transportation / travel plan 	<ul style="list-style-type: none"> • Conduct / Receive training • Install and practice software • Input sample data and check quality • De-bug and eliminate viruses from computers 	<ul style="list-style-type: none"> • Conduct training and receive if necessary • Complete sampling plan and assign relevant tasks to Teams 1 & 2 • Make sure that the logistics, materials and budget are in order • Coordinate with Team 1 to translate, test and adjust survey
Phase Two: Field Work	<ul style="list-style-type: none"> • Collect data • Follow established quality control procedures • Maintain schedule and coordinate adjustments with Team 3 • Maintain feedback loop on survey and procedural issues 	<ul style="list-style-type: none"> • Input and clean data • Return incomplete interview forms back to Team 1 • Maintain an accurate count of the sample accomplished each day • Ensure that survey tracking and storage procedures are followed • Maintain digital backups of data 	<ul style="list-style-type: none"> • Review quality of data collection, data cleaning and input • Ensure sampling plan and timeline are being followed • Ensure that survey tracking and storage procedures are being followed • Maintain overall flow and communication of project • Maintain budget
Phase Three: Analysis and Project Close-out	<ul style="list-style-type: none"> • Finalize data collection if necessary • Conduct debriefing of Interviewers before they are dismissed • Supervisors help with data cleaning if needed • Write field report(s) (if applicable) 	<ul style="list-style-type: none"> • Complete data input • Clean data and ensure quality control • Conduct poverty calculation and other analyses as applicable 	<ul style="list-style-type: none"> • Supervise final data collection, cleaning, input, and analysis • Help coordinate report writing (if applicable) • Send PAT results to MRR, IRIS and etc

See Appendix A for more information.

Logistics

Well-planned logistical support—coordinated transportation, communications, field supplies, and contingency plans for disruptions—also greatly enhances the quality of field implementation. Logistical support needs to be carefully planned at all stages of the survey process, especially where operations take place in remote locations with limited infrastructure.

Some important logistical considerations:

- Consider local customs, holidays and political circumstances when deciding on the time of year to implement.
- Work with local field staff to develop realistic interview schedule, including planning for local and long distance travel and logistics and scheduling interviews at times when respondents are most likely to be available
- Route planning should include access to petrol stations, food, and accommodations, if necessary.
- Communication methods and back-up plans also should be identified (cell phones, phone cards, alternate sites to visit etc.).

Equipment

Your organization will need to dedicate the appropriate resources to the poverty assessment to ensure its success. The requirements for each implementation will be different but they should include:

Transportation: The vehicles and drivers necessary for the implementation team to conduct the field work. Vehicles should be large enough to carry the field team and supplies, and sturdy enough to withstand road conditions in survey areas. (It might be more cost-effective to use taxis or other public transport in urban areas.)

Computers: When possible, computers should be designated full-time for use **only** by the poverty assessment team during the period of field work and data entry. Consider report writing (if needed), scheduling, creating training materials, and data entry. With two Data Processors, there should be two computers exclusively for data entry, if possible. The exact number should be determined by the Project Manager during planning.

Photocopying: Several hundred copies of the survey form will need to be copied, collated and stapled. Your organization should consider outsourcing this task, as it can be time-consuming. If possible, the surveys should be printed two-sided, to minimize the need for stapling, to reduce paper costs, and reduce the chance of separating or losing pages.

Office Space: Office space and desks should be dedicated to the poverty assessment team during the field work and data entry activities. Adequate training and working space will be necessary throughout the planning and implementation processes. The data processing should be done in a separate room to avoid distractions to the personnel.

Workload

In some instances, it may make sense to combine some of the recommended implementation team roles. If reducing the proposed team size, it is important to remember how combining roles might affect quality control and whether or not it might ever require someone to be in two places at once. For example, if the SSTC is used as a Field Supervisor, they may not be available when another Field Supervisor needs help making an adjustment to their interviewer schedule for the day. Also, that individual is then responsible for both turning in and properly storing their own survey; there is no one to monitor the process and provide an extra layer of quality control (making sure the surveys are complete and being put in the right place).

Timeliness is another factor to consider. Creating a detailed chart of proposed activities and their timing may help you see whether or not someone has been ‘double booked’ and is needed to serve too many functions at once. See the following example. When considering timing, also remember that any number of activities may take longer than anticipated. Make sure you have ‘back ups’ prepared if a team member is unable to perform their responsibilities because the project ran longer than expected.

PAT Survey – Pre-implementation planning and logistics			
Role	Name	Time-on-task (total this phase only, days)	Within Normal Scope Of Work?
Project Manager	Vladimir Terskin	9	Yes
Logistics Assistant	Desta Amde	3	Yes
Sampling & Survey Tracking Coordinator	Tracy Choi	4	No
Field Supervisors	TBD	6	No
Data Processing Coordinator	Elliott Skabler	5	No
Data Processors	TBD	2	No
Interviewers	TBD	3	No
Tasks		Time-on-task (by task)	
Hire implementation staff, as needed	V.T.	2	
	D.A.	1	
Train core staff (Field Supervisors, Sampling & Survey Tracking Coordinator, Data Processing Coordinator)	V.T	5	
	D.A.	2	
	T.C.	3	
	Field Supervisors	3	
Train Interviewers	E.S.	3	
	V.T.	2	
	Field Supervisors	3	
Train Data Processors	Interviewers	3	
	T.C.	1	
	E.S.	2	
	Data Processors	2	

Timeline

	June 1-7	June 8-14	June 15-21	June 22-28	June 28-5	July 6-13	July 14-20
Hire staff	■						
Train core staff		■					
Train Interviewers			■	■			
Train Data Processors			■	■			

CHAPTER THREE – SAMPLING

Sampling is what we do when we want to find out information about a large group of people but cannot or do not want to gather information on every single person (called a census). A census can be costly, extremely time consuming, and because it requires intensive supervision and monitoring, is prone to error.

The techniques or methods you use to choose your sample is called your *sample design*. A sample design involves:

- Identifying your target population (*sampling universe – who do you want to learn about*)
- Selecting a sampling methodology, including
 - Method of sample size calculation
 - Method of sample selection

In this section, we will provide an overview of three commonly used sampling methodologies – **Simple Random Sampling, Cluster Sampling, and Stratified Sampling** – that you may employ when setting up your sampling frame – the list(s) that identify your target population – and calculating the sample size you will need to maximize the accuracy of your findings. We will also describe several methods of random selection for selecting the clients / beneficiaries to be included in your sample. Sampling is heavily dependent on context. The information presented in this chapter will be focused on the theories and principles that should guide your sample design, keeping in mind that your organization will have to make adjustments to meet your specific needs and goals. It would be impossible to suggest one or even several specific sample designs that would work for most PAT implementers. Use the guidelines set out in this chapter to guide your sample design and consult with the PAT Help Desk as needed.

Oversampling

Clients who are unavailable or unwilling to participate, surveys that get damaged or lost, interviewers who do not follow quality control procedures — these are just some of the reasons you might need to interview more clients than indicated by your calculated sample size. With all of the forthcoming examples, we recommend you oversample: increase your sample size, select and interview more clients than your sample size requires.

To oversample, first, calculate your required sample size, then decide what additional percent of clients you think you should select. Consider your anticipated response rate and a reasonable degree of error. When possible, consult with others who have previously implemented surveys for your organization. MFIs using the PAT have reported oversampling up to 65% but you may not need that much.

Whether to interview extra clients during regular field work or to wait until your selected sample is exhausted to conduct further interviews is at your discretion. Interviewing extra clients within your regular implementation schedule takes time and thus costs more money, but it may cost considerably less than it would cost to return to the field later and is thus recommended.

Target population – Who do you want to learn about?

It is essential that you specifically define the group that you want to study. This is called your **target population** or **sampling universe**. For example, you may want to look at the prevalence of poverty among all of the clients of your organization, or you may want to focus only on the new clients that have been with your organization for six months or less.

If you are required to use a PAT as part of USAID reporting, you have the option of sampling from all current clients, or just from “new” clients (e.g., clients who have joined in the last six months). If you are going to sample new clients or beneficiaries, be sure they do not think that their answers to the PAT questionnaire will affect their program participation, loan size, etc. To limit this effect, you should give PAT surveys only after the person knows they have been accepted to the program or has received their first disbursement.

When choosing your target population, consider how easily you will be able to obtain a complete list of all clients / beneficiaries in that population. These records and the information they provide about each member of your target population make up your **sampling frame**. The level of detail and completeness of your sampling frame will have implications for the sample design you can and should use. For example, if you are conducting in-home interviews, you will need to know where everyone lives. If you are conducting meetings in your office or at another central location, you will not need a complete client list but you will need to know the number of

clients and when they are likely to be at a particular location⁸. If one of your branches offers group loans and cannot provide you with a list of names, only descriptions and locations of group meetings, you will have to adopt a design that allows for that (such as a cluster design).

If you are having difficulty compiling a complete sampling frame, you might want to consider defining your target population in a different way.

Sampling methodology⁹ – Simple Random Sampling (SRS)

Essentially, SRS is when you choose a sample of your clients/beneficiaries at random from your entire target population. Using this method, each and every client will have an equal probability of being selected into your sample. One of the most difficult things about SRS is meeting the criteria for using it.

Practically speaking, you can choose your sample using SRS if you:

1. Have a complete client list, including
 - Names and addresses if you're meeting clients in their homes OR
 - Knowledge of when and where all of your clients will be accessible¹⁰ if you are conducting surveys at your branch or other offices.
2. Can realistically reach any random sample of clients, however spread out the sample may be. If your clients are spread out over a very large area, and/or some are very difficult to reach, it may not be realistic to select from the entire list of clients, as it would take too much time and money to travel to every client in the sample.

If you cannot meet both of these requirements, you will not be able to use SRS as an accurate sampling methodology.

⁸ More specifically, you will need to know the time period in which all clients will have to appear in a particular place (or at one of several places). For example, if loans are disbursed during the first week of every month, you would have your interviewing team conduct interviews during that week.

⁹ Your sampling methodology will have implications for your sample size, sample selection, and to some degree, the accuracy of your findings. This is one of the most critical phases of sample design and we highly recommend consulting with a sampling expert to determine the best methodology for your situation. The advice presented here is general; sampling is highly dependent on the unique characteristics of your target population, the type of identifying information you have on them, geographic and linguistic factors, and resource availability.

¹⁰ Note that being accessible and appearing at a particular location are different. You might know when all members of a group loan will be meeting but they might not have time to be interviewed. It is always advisable to let people know ahead of time that you would like some of their time for an interview. One option is to use group loan meetings as an opportunity to set appointments to meet individual members at their homes at another time.

Calculating the Sample Size

If you have decided that you will be able to meet the criteria for SRS, you are ready to **calculate your sample size**. There are a number of online SRS sample size calculators. We are going to show you how to use one in particular¹¹. Follow the guidelines below to use the calculator.

How much error are you willing to tolerate?
5% error is a common statistical convention though you should use your own discretion to choose the error level.

How many people are in your population?
Enter the total number of clients in your target population.

Press 'Calculate.' You should use the results for a Confidence Level of 95% (another common, statistical convention).

Once you have calculated your sample size, you are ready to select your sample. We will discuss several methods for doing so later in this chapter.

USAID Reporting Requirement

USAID stipulates that a sample size of 300 is sufficient for meeting its reporting requirement. This is because, as you can see from the chart below, assuming a confidence level of 95% and 5% or 5.5% error, a sample size of 300 (or slightly more) is sufficient regardless of population size. However, this is assuming you are using Simple Random Sampling. Understand that if you employ a different sampling methodology, you will likely assume a greater degree of sampling error in your results using a sample size of just 300.

Population	Sample Size – 5% error	Sample Size – 5.5% error
500	217	194
700	248	218
800	260	227
900	269	235
1,000	278	241
2,000	322	274
3,000	341	287
4,000	350	294
5,000	357	299
10,000	370	308
20,000	377	313
50,000	381	315
100,000	383	316

¹¹ www.custominsight.com/articles/random-sample-calculator.asp

Sampling methodology – Cluster Sampling

If your client population is spread out over a large geographic area, it may be difficult and time-consuming to visit every potential location in which randomly selected clients can reside. In this case, you might want to use existing geographic groupings among which your client population is dispersed and then chose random clients among only certain groups. These geographic groupings ('clusters') will be naturally formed, such as around a branch or regional office. This can be beneficial when your client population is spread out over a large geographic area and reaching clients across the entire region could be too costly or time-consuming. **Cluster Sampling** will limit the number of regions or clusters your interviewing team will have to visit. After clusters are identified, you will select a pre-calculated number of clusters to be included in your sample and then select a pre-calculated number of clients from each cluster to form your sample.

Accurate cluster sampling requires you to have and/or calculate several pieces of information.

What are your clusters? As previously mentioned, your clusters should be composed of existing geographic groupings within your organization. For example, the clients served by each branch office could compose your clusters. Ideally, clusters should be relatively small. It is likely that separating clients by branch office into 30 clusters will yield a better design than, for example, separating them by county into 6 clusters. While there is no specific optimal cluster size, you will want to balance the fact that the larger the clusters, the larger your sample size will be and the smaller the clusters, the more dispersed your sampled clusters are likely to be¹².

What is the rate of homogeneity (roh or ρ) for your population¹³? In our case, roh is the likelihood that two clients or beneficiaries who are in the same cluster will have the same level of poverty, as compared to two clients or beneficiaries who were selected at random from the entire population. For example, a roh of 0.02 means that clients or beneficiaries within a cluster are 2% more likely to have the same poverty level (to be similar, or, homogenous) than if they'd been chosen at random from the entire population.

There are several options for finding a value of roh. First, look to results of previous studies done by your organization or other local organizations. A roh calculated from another study even on a different topic can be used if the study topics are analogous. For example, if it is known (through research) that food security and poverty are closely correlated in your region, then using a value of roh calculated from a recent food security survey can be used.

Other sources can include your local or national statistics office or results from larger national surveys, such as the USAID-funded Demographic and Health Surveys (www.measuredhs.com). Contact the PAT Help Desk (pathelp@iris.umd.edu) if you need assistance finding a value of roh.

¹² Imagine clusters of size 1. This would be the same as SRS and would yield none of the benefits of cluster sampling (as would be the case with other very small cluster sizes). Next imagine dividing your population into three large clusters. Your sample size would be large enough to require you to select all three clusters, again negating the benefits of cluster sampling.

¹³ Roh also called coefficient of intraclass correlation.

What is the average per cluster cost of conducting the PAT? Costs will vary from cluster to cluster because of increased travel and other costs so the average must be calculated. Costs to consider: costs of travel to the region, costs of travel around the region, car rental, fuel, bus, taxi, lodging, meals, etc. *This is referred to as c_a below.*

What is the average cost per interview? Consider the wage, per diem, and other costs incurred for one interviewer to interview one client. It may be simpler to calculate the cost per day per interviewer and divide by the average number of interviews you think one person can conduct in one day. *This is referred to as c_b below.*

What is the optimal number of clients/beneficiaries to interview in each cluster (b)? This calculation is based on your roh and the cost calculations (c_a and c_b) described above. It assumes you will select the same number of clients in each sampled cluster¹⁴. You will later use this calculation to determine the number of clusters to select. This is referred to as b below.

$$b = \sqrt{\frac{c_a * (1 - \rho)}{c_b * \rho}}$$

What is your design effect? The design effect (DEFF) is the *effect* that your particular sample *design* (i.e. cluster sampling) has on your sample size calculation. It is actually the factor by which your sample size – if you were using SRS – must be multiplied to get the accurate sample size to use for a cluster design.

$$DEFF = 1 + (b - 1) * \rho.$$

What is sample size required for an SRS design with your population size? As previously mentioned, you will multiply this value by your DEFF to get the sample size needed for your cluster sampling.

Cluster Sampling: Step by Step

Here are the steps you will follow to calculate the minimum sample size, number of clusters and number of clients per cluster to select for a cluster design:

- 1) Identify your clusters
- 2) Find your roh (rate of homogeneity or ρ)
- 3) Estimate the average per cluster cost (c_a) and the average cost per interview (c_b)
- 4) Calculate the optimal number of clients/beneficiaries to interview in each cluster (b)
- 5) Compute your design effect (DEFF)

¹⁴ It is possible to sample a proportionate number of clients from selected clusters (so that larger clusters have larger sample sizes) or even to select all clients from each sampled cluster. These methods will require slightly different sample design procedures which we will not cover in this manual.

- 6) Calculate the SRS sample size for your population size with an online calculator¹⁵
- 7) Calculate the sample size for your cluster design
- 8) Calculate the number of clusters to select for your sample

EXAMPLE: The Mingunesian Microfinance Association¹⁶

Recall that the MMA serves 2,300 clients spread over 45 branch locations with three programs. In order to conduct their PAT, they've decided to use a cluster design. Their sampling universe includes all 2,300 clients. See below how they used the process described above to calculate their sample size and select clusters and clients to be included in the survey.

1) Identify your clusters

The MMA has 45 loan groups so each group will serve as a cluster.

2) Find your roh (rate of homogeneity or ρ)

A previous survey identified a roh of 0.02.

(For steps 3 – 8, see a copy of the MMA's Excel spreadsheet and calculations on the page following the notes below.)

3) Estimate the average per cluster cost (c_a) and the average cost per interview (c_b)

The MMA's loan groups (clusters) are located in 4 separate regions. Costs for traveling to these regions were calculated and an average was taken, based on the total number of clusters per region.

Costs considered for the 'per cluster' (c_a) calculation included:

- Cost and amount of gas used for travel to region (three regions)
- Cost and amount of gas used for travel around region (three regions)
- Car rental (three regions)
- Bus travel (for one region)

These calculations yielded $c_a = \text{MNP}^{17} 200.00$.

The cost of a single interview was calculated by estimating the "cost" of hiring an interviewer for a day and dividing it by the average number of interviews one could complete in a day. Costs considered for the 'per interview' (c_b) calculation included:

- Interviewer wage
- Per diem (food and lodging)
@ 8 interviews per day

¹⁵ Such as the one found here: www.openepi.com/Menu/OpenEpiMenu.htm

¹⁶ While this is a fictitious example, it is based on the plan that IRIS helped an actual PAT implementer develop.

¹⁷ Recall that MNP = Mingunesian Pounds, the local currency in our example.

These calculations yielded $c_b = \text{MNP } 25.00$.

4) Calculate the optimal number of clients/beneficiaries to interview in each cluster (b)

$$b = \sqrt{\frac{c_a * (1 - \rho)}{c_b}} \quad b = \sqrt{\frac{\text{MNP}200}{\text{MNP}25} * \frac{1 - 0.02}{0.02}} \quad b = 19.79898987(\text{round to } \mathbf{b=20})$$

5) Compute your design effect (DEFF)

$$\text{DEFF} = 1 + (b - 1) * \rho$$

$$\text{DEFF} = 1 + (20 - 1) * 0.02$$

$$\text{DEFF} = 1.38$$

6) Calculate the SRS sample size for your population size with an online calculator¹⁸

The MMA specified they would tolerate +/- 5% error with a 95% confidence level.

Using the online calculator, they got an **SRS sample size of 329**.

Talent Management Solutions

CustomInsight
.com

[Products/Services](#)
[Samples](#)
[Insight](#)
[Clients](#)
[Help/FAQ](#)
[Contact](#)

[360 Degree Feedback Topics](#)
[Employee Survey Topics](#)

Survey Random Sample Calculator

If you have not yet conducted your survey and want to calculate how many people you need for your random sample, use calculators #1 and #2 below.
If you already conducted your survey and want to know how accurate your data are given the number of responses you received, scroll down and use calculator #3.

#1 - How many survey respondents do you need?

Specify your desired error level and population size below and click calculate. The numbers next to each confidence level indicate how many people need to complete your survey to achieve the specified error level.

How much error are you willing to tolerate?
If you are not sure, try somewhere between 3% and 6%

 %

90% Confidence	243
95% Confidence	329
99% Confidence	515

¹⁸ Such as the one found here: www.custominsight.com/articles/random-sample-calculator.asp

7) *Calculate the sample size for your cluster design*

Cluster design sample size = DEFF * SRS sample size

Cluster design sample size = 1.38 * 329

Cluster design sample size = 454

8) *Calculate the number of clusters to select for your sample*

Number of clusters to select for sample = Sample size / b ¹⁹

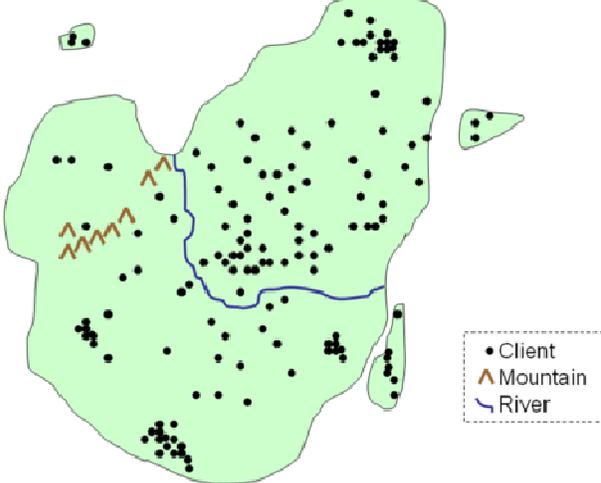
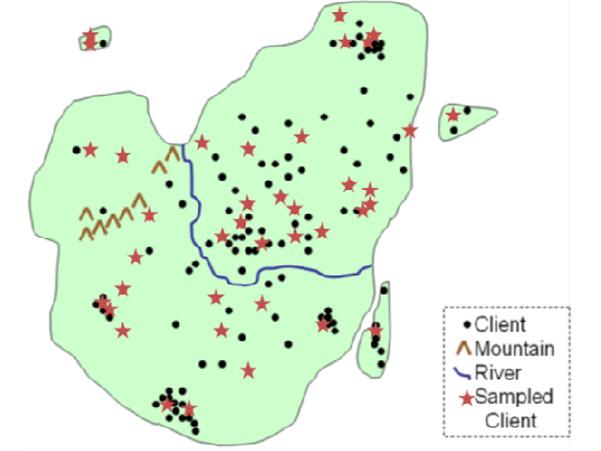
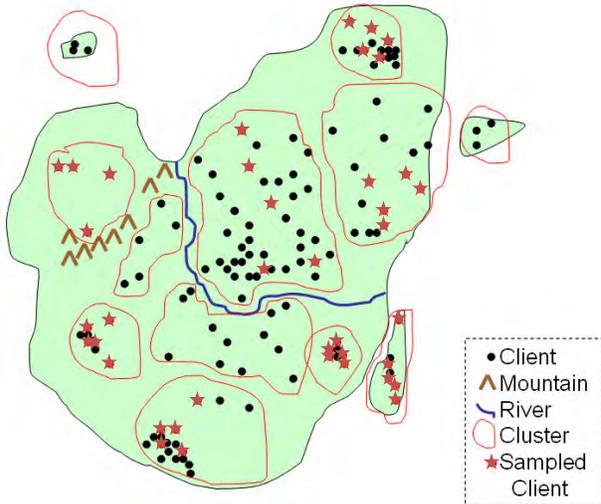
Number of clusters to select for sample = 454 / 20

Number of clusters to select for sample = 23

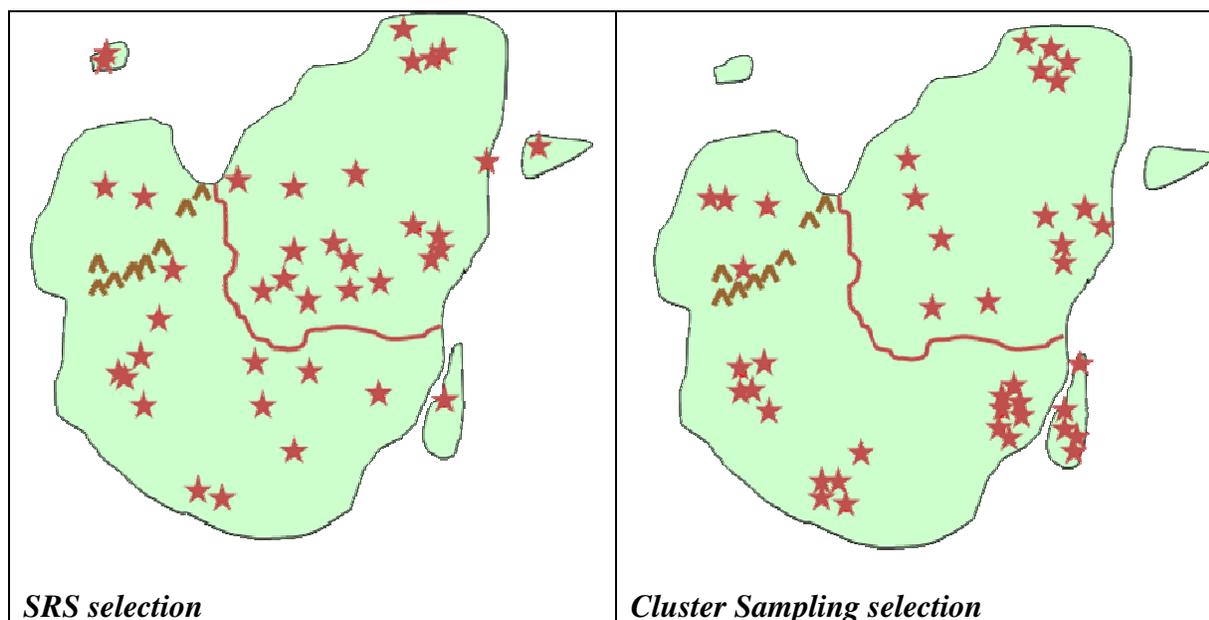
¹⁹ Recall that b is the optimal number of clients/beneficiaries to interview in each cluster calculated in Step 4.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	CLUSTER DESIGN EXAMPLE													
2	* Currency is expressed in MIngunesian Pounds (MNP)													
3	c_a calculation					gas @ MNP 4.00 per gallon * 17 gallons per tank		MNP 68.00 per tank						
4	--- average cost per cluster													
5		Region	# of clusters (loan groups) per region	proportion of day to travel to site (R/T)	days to spend	gas @ 1 tank per day (travel to site)	local travel gas per day	total local gas	car rental @ MNP 40.00 / day * 2 cars	bus cost: 4 enumerators @ MNP 5.00 / day	total bus	total cost / cluster	* # of clusters	
6		B	15	0.1	4					MNP 20.00	MNP 80.00	MNP 80.00	MNP 1200.00	
7		A	10	0.25	4	MNP 17.00	MNP 34.00	MNP 136.00	MNP 80.00			MNP 233.00	MNP 2330.00	
8		C	9	0.5	4	MNP 34.00	MNP 34.00	MNP 136.00	MNP 80.00			MNP 250.00	MNP 2250.00	
9		D	11	1	4	MNP 68.00	MNP 34.00	MNP 136.00	MNP 80.00			MNP 284.00	MNP 3124.00	
10		# clusters=		45								average:	MNP 197.87	
11												rounded:	MNP 200.00 =c_a	
12	c_b calculation													
13	--- average cost per interview													
14														
15	Enumerator's wage per diem	MNP 150.00 per day												
16	per diem	MNP 50.00 per day												
17	interviews	8 per day												
18														
19	Cost per interview=	MNP 25.00 = c_b	FORMULA: =(B15+B16)/B17											
20														
21	roh =	0.02	KNOWN											
22														
23	calculate b:													
24	--- optimal # of clients to sample per cluster													
25														
26	b =	20	FORMULA: =ROUND(POWER((M11/B19)*((1-B21)/B21),0.5),0)											
27														
28														
29	calculate DEFF:						calculate # of clusters to select							
30	--- the effect your sample design has on your sample size													
31														
32	DEFF =	1.38	FORMULA: =1+((B26-1)*B21)											
33														
34	calculate sample size for population using SRS													
35														
36	sample size for population of 2300 using SRS=	329	FROM SAMPLE SIZE CALCULATOR											
37														
38	calculate sample size for cluster design													
39														
40	sample size =	454	FORMULA: =ROUND(B32*B36,0)											

SRS vs. Cluster Sampling – an example in pictures

<p>Here is a country map with all of an organization's clients represented by black dots. Notice that the clients are spread across the country, including on each of three smaller islands.</p>	 <p>Population map</p> <ul style="list-style-type: none"> • Client ▲ Mountain — River
<p>This map shows the results of SRS. The sampled clients are represented by red stars.</p>	 <p>SRS selection</p> <ul style="list-style-type: none"> • Client ▲ Mountain — River ★ Sampled Client
<p>This map shows the results of Cluster Sampling. The population has been divided into geographic clusters, several clusters were chosen, and several clients within each cluster were chosen.</p>	 <p>Cluster Sampling selection</p> <ul style="list-style-type: none"> • Client ▲ Mountain — River ○ Cluster ★ Sampled Client

The impact of cluster sampling is made clearer when we take the lines and non-sampled clients away and compare the two maps:



As you can see, when Cluster Sampling is used, the sampled clients are not spread out as far as when using SRS, even with a larger sample size. Also, notice that only one island must be accessed in the cluster sampling example.

Stratified Sampling

Stratification is the process of partitioning a population into specific groups. For example, one might define two strata of male and female clients. Stratified sampling involves using proportional sample selection to ensure that each stratum is adequately represented in the sample. If your population is 20% male and 80% female and you want to make sure that men are adequately represented in our sample, you could calculate your sample size and select 20% of it from only male clients and 80% from only females²⁰. Stratification can also be done at the cluster level. For example, you might divide your clusters into rural and urban and select clusters proportionally.

Stratification can be built into an SRS or cluster sampling design. In either case, the procedures described previously for these designs must be used to calculate the appropriate sample size.

The differentiation is in the arrangement of cluster and/or client (beneficiary) lists. Interval selection (to be described in the next section) must be used and cluster and client lists must be ordered such that members of the same strata are grouped together.

²⁰ It should be noted that gains from stratifying at the client level are minimal. In theory, random client selection should result in sample selection that mirrors proportions in the population. If your population is 75% rural, your sample will likely be close to 75% rural. In other words, stratification at the client level may not be worth the effort.

Sample Selection

In this section, we will describe three methods of sample selection:

1. Using a random number table
2. Ordering lists by random number assignment
3. Interval selection
 - Cluster level
 - Client level

Using a random number table

This method is useful for SRS, for selecting clients from a selected cluster in cluster sampling, or for selecting clients from a stratum in stratified sampling.

STEP 1: Get a complete client list and number each client: 1, 2, 3, etc.

STEP 2: Create a random number table that selects from numbers with a range of 1 to your total population size. The total number of figures generated in your random number table should include 20-30% more elements than the number of clients you will be selecting.

For example, if you need a total sample size of 600, create a random number table with ~750 elements. This is because random number generators will generate repeat numbers (and you cannot select an individual client more than once).

There are a number of random number generators available online²¹ or through other statistical programs. If you are creating a random number table in Excel, we recommend using one of the following functions²², where n is your total population size.

=RANDBETWEEN(1, n) OR
=ROUND(RAND()*(n -1)+1,0)

STEP 3: To prevent random numbers from regenerating (as they will do any time you modify the spreadsheet), COPY the table of random numbers, then select PASTE SPECIAL → VALUES or simply PASTE VALUES, depending on your version of Excel. What this does is replace each cell that contains a RAND formula with a random value.

STEP 4: Match the numbers on your random number table with the numbered clients in your list. When you come across a duplicate number in the table, skip it and move on to the next. Continue to select numbers until you have met your sample size.

²¹ www.openepi.com/Random/Random.htm is one.

²² RANDBETWEEN is not included in some versions of Excel which is why we offer the ROUND(RAND) function as well.

Ordering lists by random number assignment

This can also be done at the client level. It does NOT work for stratified sampling.

STEP 1: Create a client list in Excel or a similar program.

STEP 2: For a population of size n , assign a random number to each client (or cluster). We recommend one of these two formulas²³:

=RANDBETWEEN(1, n) OR
=ROUND(RAND()*($n-1$)+1,0)

STEP 3: To prevent random numbers from regenerating (as they will do any time you modify the spreadsheet), COPY the list of random numbers, then select PASTE SPECIAL → VALUES or simply PASTE VALUES, depending on your version of Excel. What this does is replace each cell that contains a RAND formula with a random value.

STEP 4: Sort the list by the random number column. As noted before, a single random number may be assigned to more than one item in your list. This is expected and has no effect on the ‘randomness’ of your list.

STEP 5: Number the list sequentially, either in a separate column or over top of the random number column. The list should order clients: 1, 2, 3, 4, 5, etc.

STEP 6: For sample size m , select clients numbered 1 through m to be included in the sample.

²³ RANDBETWEEN is not included in some versions of Excel which is why we offer the ROUND(RAND) function as well.

Interval selection at the cluster level (Systematic Probability Proportion to Size (PPS) Sampling)

This method of sample selection ensures that clusters are chosen in proportion to their size. Thus a cluster of 500 clients will have a greater probability of being selected than a cluster of 200 clients. However, this methodology ensures that each client has an equal probability of being selected.

Assume we are to survey a sample size of 30 from the following population by selecting 15 clients from each of 2 clusters:

Cluster	# Clients
Cluster A	25
Cluster B	20
Cluster C	15
Cluster D	25
Cluster E	15
Total	100

We want each client to have an equal chance of being selected into the population. A client's probability of selection = the probability of selecting that client's cluster multiplied by the probability of selecting that client from the cluster.²⁴

$$P_{(\text{client in A})} = P_{(A)} * P_{(\text{being selected in A})} = 2*(25/100) * (15/25) = (1/2) * (3/5) = 3/10 = \mathbf{0.30}$$

Note that P(A) is multiplied by 2 because we are selecting 2 clusters out of the 5. Now compare:

$$P_{(\text{client in B})} = 2*(20/100) * (15/20) = (2/5) * (3/4) = 6/20 = \mathbf{0.30}$$

$$P_{(\text{client in C})} = 2*(15/100) * (15/15) = (3/10) * (1) = 3/10 = \mathbf{0.30}$$

$$P_{(\text{client in D})} = 2*(25/100) * (15/25) = (1/2) * (3/5) = 3/10 = \mathbf{0.30}$$

$$P_{(\text{client in E})} = 2*(15/100) * (15/15) = (3/10) * (1) = 3/10 = \mathbf{0.30}$$

Each client has a 0.30 or 30% chance of being selected into the sample.

Setting up this kind of sample selection is straightforward.

Once you've decided how your clusters will be divided and how many you need to choose, make a chart that lists each cluster and its size. If using strata, arrange the list by stratum. Then designate an interval for each cluster based on its size and the cumulative population total.

For example, assume Cluster A contains 312 clients and Cluster B contains 426 clients. Cluster A's interval is 1 - 312 and Cluster B's interval is 313 – 738 (for an interval size of 426, equal to the number of clients in the cluster). An individual cluster's interval includes all numbers between the previously listed cluster's maximum and its listed maximum (inclusive of the latter).

²⁴ This is often referred to as two-stage sampling or two-stage cluster sampling.

We will use an example to illustrate the interval calculation and cluster selection process. The following chart lists the clusters, number of clients and interval for a sampling universe.

Cluster	# Clients	Interval
Cluster A	312	1-312
Cluster B	426	313-738
Cluster C	396	739-1134
Cluster D	434	1135-1568
Cluster E	392	1569-1960
Cluster F	401	1961-2361
Cluster G	512	2362-2873
Cluster H	489	2874-3362
Cluster I	412	3363-3774
Total	3774	

As previously described, the reason we use intervals is so that clusters with a larger population have a greater chance of being selected than those that are smaller. In order to select the actual clusters, calculate an interval of selection: $i = \text{population total} / \text{number of clusters to select}$. Then pick a random number between 1 and interval i . Then, beginning with the cluster that includes i in its interval, select every i th cluster to be included in the sample.

In the example above, assume 4 clusters need to be selected for the sample. i is calculated to be $3774/4 = 944$. Suppose we pick random number 334 to start with. That means our first selected cluster is Cluster B because 334 is within its interval of 313 – 738. We then add our interval to our random number ($334 + 944 = 1278$) to get our second cluster selection (Cluster D) and so on. If adding i to a previous selection results in ‘landing’ in the same cluster again, you will simply take 2 samples from that selection²⁵.

Example

		Selection
Start:	334 (random #)	<i>Cluster B</i>
Next:	$334 + 944 = 1278$	<i>Cluster D</i>
Next:	$1278 + 944 = 2222$	<i>Cluster F</i>
Next:	$1002 + 944 = 3166$	<i>Cluster H</i>

Now we have our 4 selected clusters.

²⁵ In reality, you would take one sample that is twice as large as your per-cluster sample size (b).

Interval selection at the client/beneficiary level (Systematic Random Sampling)

Selecting clients is much like selecting the clusters, only a bit simpler.

STEP 1: Create a client list.

STEP 2: Number the clients (1, 2, 3 etc).

STEP 3: Calculate your interval of selection: $i = \text{total \# of clients} / \text{number of clients to needed to fill sample size requirements}$

STEP 4: Select a random number between 1 and i

STEP 5: Starting with the random number, select every i th client to be included in your sample, until you reach sufficient size.

EXAMPLE: Your first client selected, is the one whose assigned number is equal to the random number you selected. The second client selected is number i plus the first client's number. The third client is i plus the second client's number and so on. Here is an example of a list of the *selected* clients where RN is the random number selected:

#	Cluster A	Client Since	Address
RN	A Wooks	2007	8010 Cherrywood Dr
RN+ i	G Morguson	2008	900 Gunners Ct
RN+2 i	A Sulliago	2009	1135 S Ninth St
RN+3 i	B Chen	2009	3205 Portia St, Apt 1A
RN+4 i	T Doughez	2008	1901 W O St
RN+5 i	R Craver	2008	2900 S 27th St
RN+6 i	R Millack	2009	2521 Cindy Dr
RN+7 i	D Ndella	2007	5230 Heumann Dr
	...		
	...		

Note for those conducting interviews in an office or other location outside the household. If it is impossible to get access to client lists, you might consider an ad hoc interval selection at a branch or loan disbursement office. Calculate an interval of selection and then select every i th person that enters the office or stands in line, etc. Be sure to consider carefully any bias in this methodology. For example, within your sampling universe, are there some kinds of people that are more likely to come at a particular time or on a particular day? Are there some people that never come to that particular office at all?

CHAPTER FOUR – PLANNING FOR FIELD WORK

Preparing the PAT Survey

Whether you use the PAT survey in English or translate it into one or more languages, it is essential that you stick as closely as possible to the original survey. That includes adhering to the original wording of questions (as closely as possible if translating), maintaining the order of the survey, not adding questions to the beginning or body of the survey²⁶, and not omitting any questions. Survey questions have been phrased, coded, and put in a specific order and format to produce consistent, measurable results. Altering the survey in any significant way could:

- Introduce bias into the survey, or
- Invalidate survey results

Translation

In most instances, the PAT will need to be translated from English into one or more local languages. The entire questionnaire must be translated before the interviewer training though adjustments should be made based on feedback from Interviewers during practice interviews. Allowing the Interviewers to translate during the interviews will bias the results, since each interviewer might phrase a question differently. The only instance where a written translation into the local language is not required is when the local language does not have a written form.

An early step in the translation process should be a discussion by the implementation team leaders on how to translate the survey. Staff members who interact with the clients on a regular basis should be included in this discussion to make sure that the local terms being used in the translation will be understood easily by those being interviewed. Once everyone agrees on a ‘dictionary’ of terms to be used, the tool can be fully translated and available in printed form in the local language for the interviewing process.

Another essential step in the editing of the translation is to “back-translate” into English by someone who has not seen the original English version. A comparison of the back-translation and the original English version is the best way to ensure that the translation is similar to the English original. The differences can assist in editing and improving the translation.

Bias may creep into survey items in a variety of ways when translating. Consider the following when customizing the questionnaire and training your staff.

- *Leading questions.* A leading question is one that seems to lead or encourage the respondent to answer in a certain way. For example, “You like the way this manual is written don’t you?” makes it harder for a person to say that they do not like the manual. In contrast, “Do you like the way this manual is written, yes or no?” does not lead the respondent. When you translate questions it is important to maintain the intended structure and not introduce or rephrase the question so that it suggests a specific response.

²⁶ It is okay to add questions to the **end** of a PAT survey but they will NOT be included in the poverty calculation.

- *Unfamiliar terms and jargon.* Wherever possible, use words most likely to be understood by the respondent while retaining original meaning.
- *Poor grammatical format.* Poorly worded sentences can introduce bias by confusing the respondent.

The translation of the interviewer's introduction should be adapted to your local context, taking into account cultural norms and expectations for how one should be greeted. Your introduction should include the following components:

1. **Introduction:** Clients will be more cooperative if they know who is conducting the study. An important point to mention is that the survey team is not directly involved in making decisions on the clients' participation in or access to services. In some cases, a letter of introduction from your headquarters to clients and from local authorities to non-clients can reassure clients and further facilitate introductions.
2. **Purpose:** It is recommended those being interviewed not be told that the ultimate focus of the survey is poverty measurement. This information could influence the way that questions are answered by the clients and thereby introduce a major source of error in the results. Letting them know that their participation will help the organization provide better services and products should be sufficient.
3. **Time:** Let clients know that the survey is brief and will only take X minutes of their time.
4. **Confidentiality:** In many countries, fear of crime or traditional beliefs may also inhibit many people from sharing private information. Introductions should incorporate clear statements about the neutrality of the interviewer and the confidentiality of information collected for the study. The interviewer should guarantee that the respondent's specific answers will not be shared with program staff or anyone else, and that the data will only be examined in the aggregate.
5. **Consent:** Be sure to give the client the opportunity to accept or refuse the opportunity to participate in the survey. Gentle reassurance can be used to help a client agree to be surveyed but at no point should they feel obligated.

See page 45 for details on pre-testing the questionnaire for translation and other issues.

Printing & Copying

Allow adequate time for photocopying surveys, whether you are preparing the copies in your office or sending them off to be copied elsewhere. It is preferable to use double-sided copies (rather than stapled, one-sided pages) whenever possible. This eliminates the risk that the pages of a single survey will be separated. If you cannot make double-sided copies or if your survey contains more than two pages, be sure to include the Survey Number at the top of each paper in case they do become separated.

Developing a Survey Tracking System

USAID requires that you keep your original survey forms for a minimum of **THREE** years. Keeping track of your surveys on a day-to-day basis during the survey process is also important. If errors are found during the data entry or analysis processes, it should be easy to find the original survey form in order to verify what was written.

All Core Team members should be aware of the system developed for filing and storing surveys. Not only must completed surveys be easily accessible, but there must be a clear system for numbering surveys in accordance with your sampling plan.

Surveys should be stored in a secure area with minimal risk of being damaged. Ideally, a water-proof and fire-proof file or storage cabinet should be used. At minimum, papers should be in a place free from risks of being spilled on, blown away, torn, or damaged in any way.

Scheduling Interviews

The poverty assessment interviews should only be done **AFTER** the client has been accepted to receive a loan or other product or service, which will lower the temptation for the client to say what they think the institution wants to hear. If your organization accepts new clients or beneficiaries only at certain times of the year, this might affect the timing of your surveying—especially if you choose to sample only from new clients.

You should also consider times of the year and the time of day when people are likely to be particularly busy or more likely to have time available to be interviewed. Planting and harvesting seasons, rainy seasons, or other annual occurrences can be bad times to try to access people.

The interview itself may not take that much time to conduct. However, locating the correct client, making introductions, and departing smoothly can easily triple the time needed. Depending on how far apart the interviewees are, the transportation will usually take much more time than the interviewing process. That is why it is very important to solicit help from the local staff who know where their clients live and when they are most likely to be working and unavailable. A great deal of travel time can be saved by receiving adequate logistical guidance from local staff.

A minimum target of eight interviews per day for each interviewer is recommended, although some adjustment may be needed to reflect survey length and logistical conditions. If the travel time and logistics are relatively easy, up to 10-12 interviews could be done in one day. If the clients are coming to the branch office or other meeting place, as many as 15 could be done in one day. Keep in mind that too many interviews per day could compromise interview quality, while too few interviews per day could increase field costs. It is critical that each interviewer spends time immediately after each interview to make sure the survey form is complete, with answers to every question, and will be readable by the Field Supervisor and data entry personnel.

How many Interviewers will I need?

This will depend on several factors. You will need to determine your limiting factors, such as:

- Time available to conduct surveys
- Budget
- Travel required
- Approximate number of interviews per day (might be different in different regions)
- Number of Interviewers you can hire (based on cost, languages needed, resources available, etc.).

For example, assume time is your limiting factor. You must conduct 400 interviews and you only have 15 days. You have three regions to travel to (each with different numbers of people to interview) and you will bring your interview team with you from region to region. This decision was made because you realized that it costs less to transport, house and feed a team than it would to re-train new teams in each region. After pre-testing your survey, you think Interviewers will be able to conduct 8 interviews per day in any of the regions. You must also allow for 2 days of travel, meaning you really only have 13 days to conduct interviews:

Calculate: 400 interviews / 13 days = approximately 31 interviews per day

	Total interviews	Interviews per day	Approx # of days (total interviews ÷ 31)	Since we want whole days:	Interviews per day per interviewer	Interviewers needed (31 ÷ 8)
Region 1	200	÷ 31	6.5	6	8	4
Region 2	125	÷ 31	4.0	4	8	4
Region 3	75	÷ 31	2.4	3	8	4

Summary of results:

Region 1: 4 Interviewers x 8 interviews per day x 6 days = 192 interviews

Region 2: 4 Interviewers x 8 interviews per day x 4 days = 128 interviews

Region 3: 4 Interviewers x 8 interviews per day x 3 days = 96 interviews

We can see that this method of calculation is not exact (because we rounded calculations of days in each region). In the given timeline, we will not conduct enough interviews in Region 1 but have more than enough time to be in Region 3. It is best to plan ahead to compensate for this.

For example,

- Spend an extra half day in Region 1 and travel in the evening.
- Leave one of your Interviewers (with a supervisor) in Region 1 for an extra day and have one less person in Region 3 for that day – you will still be able to meet both targets.
- Consider hiring an extra interviewer or two and shortening your timeline (assuming it is cost-effective).

- Reach out to your local resources – perhaps your organization has worked with an experienced interviewer in Region 1 who could be trained quickly and work just in that region.

Note that it is always recommended to train more people to be Interviewers than the number of Interviewers required. There are always illnesses or family and work emergencies that pull trainees away at the last minute and there usually is not time to train substitutes adequately. Also, during or at the end of the training, the Core Team should have the option to recommend that certain trainees be discontinued because they did not prove during the pre-test to have the skills needed to be good Interviewers.

If you are **using loan officers or other staff members to conduct interviews**, be sure to take careful consideration of other duties they may be required to perform during the interviewing phase, and whether anyone else can cover their regular duties during the PAT implementation. Also note that using current employees as Interviewers can introduce bias into surveying, particularly if they are interviewing their own clients. The interviewer may presume answers to questions based on prior knowledge or the client may try to give answers they think their loan officer may want to hear. Thus, it is recommended that staff members do not interview anyone they already know.

Field Preparation

Before heading off to do field work, it is important to use all the resources at your disposal to learn as much as possible about the conditions and environments into which you will be travelling. This can involve issues related to local languages, customs, travel, dress, demeanor, gender dynamics and other factors.

Knowing the appropriate way to dress, greet someone, and engage in conversation are just a few things Interviewers should be clear and in agreement about. In some countries it is expected that outsiders will first seek permission from local leaders before approaching clients in a given locality. In addition to introducing the survey, these courtesy visits can also provide you with an opportunity to collect important information about the community being surveyed.

Gender dynamics are also important to consider. In some cultures it might be inappropriate for a male to attempt to enter the home of a woman when her husband is not home, or vice versa. Giving these issues consideration can impact who you hire and train to be Interviewers and how you prepare translations and implement your training.

Finally, whenever possible, give clients or beneficiaries advanced notice of when you would like to interview them. This will increase the chances that they will be available to be interviewed, and decrease the amount of time wasted trying to reach people who are not at home.

If you are **conducting interviews in your office** or somewhere else outside the interviewee's home, be sure that spaces are available to protect confidentiality during the interview. Also be sure to time interviews such that it will minimize inconvenience for the client or beneficiary. Whenever possible, give them advanced notice so that they can plan their travel accordingly.

Incentives?

Incentives such as refreshments or institutional stickers and pins can be provided for clients who participate in interviews, but do so cautiously as giving such items in exchange for a client's participation has pros and cons. It can encourage contribution but can also lead clients to believe that they 'should' give certain answers, proving that they deserve the incentive.

Providing gifts can also create a culture of expectation that could make future assessments more problematic. Additionally, giving presents to a sample of clients might create jealousy among those who do not receive anything. Stress to the client that their participation is appreciated and that their opinion/information matters. Try to convince them that the organization views them as more than "just a number."

Incentives can also be used to encourage the Interviewers to do a good job. However, if performance measures are only based on time and number of interviews, it could adversely affect quality. Thus, quality control measures such as handing in complete and legible survey forms should be included in any incentive system that is used in the field.

Training Interviewers

The Project Manager should be very much involved in the Interviewer training, working closely with the Field Supervisors or others responsible for training. Team building is important at this stage. The dialogue between those who are interviewing the clients and those who are leading the study begins here and continues with debriefing the Interviewers on a regular basis, including at the end of the interview stage.

The supervisors and all of the other people who are going to be involved in the process (Data Processors; logistics persons; etc.) should go through at least the introductory section of the same training so that they know what is being asked and expected of the Interviewers. This will allow them to be of better assistance to the Interviewers during the days of interviewing.

Several days should be devoted to training Interviewers in proper interviewing techniques, survey recording and field procedures, and general conduct. This should include ample time to conduct practice interviews and discuss lessons learned.

The following section outlines the procedures that Interviewers will need to follow, including 'best practices' to which they should adhere. You may choose to present Interviewers with handouts or use PowerPoint to share this information during training, but practicing it will ultimately be the best way for them to learn and integrate these practices into their interviews.

Before the Interview

- Be sure the survey number is recorded in the top right corner of each page. This is taken from the survey tracking or sampling list.
- Fill in as much information as possible in the top part of the survey form (see image below; other PATs will only vary slightly).

Client Assessment Survey - Bangladesh		Survey Number _____														
<i>Interviewer: Fill out the information below before the survey begins. Do not ask the respondent for this information.</i>																
Date of Interview:	<input type="text"/>	<table border="1"> <thead> <tr> <th colspan="2">Quality Control Checks</th> </tr> </thead> <tbody> <tr> <td colspan="2"><i>Field Supervisor</i></td> </tr> <tr> <td>Date _____</td> <td>Initials _____</td> </tr> <tr> <td colspan="2"><i>Headquarters</i></td> </tr> <tr> <td>Date _____</td> <td>Initials _____</td> </tr> <tr> <td colspan="2"><i>Data Processor</i></td> </tr> <tr> <td>Date _____</td> <td>Initials _____</td> </tr> </tbody> </table>	Quality Control Checks		<i>Field Supervisor</i>		Date _____	Initials _____	<i>Headquarters</i>		Date _____	Initials _____	<i>Data Processor</i>		Date _____	Initials _____
Quality Control Checks																
<i>Field Supervisor</i>																
Date _____	Initials _____															
<i>Headquarters</i>																
Date _____	Initials _____															
<i>Data Processor</i>																
Date _____	Initials _____															
Interviewer (code):	<input type="text"/>															
Branch (code):	<input type="text"/>															
Location/Region:																
1. Dhaka 3. Khulna 5. Sylhet																
2. Chittagong 4. Rajshahi 6. Barisal																
Client Location:	<input type="text"/> Urban=0; Rural=1															
Time in Program:	<input type="text"/> Months															
Client or ID #:	<input type="text"/>															

Conducting the Interview

- Begin the interview with the introduction provided. Remember, the goals of the survey introduction are to:
 - Establish whether or not the client is available to participate in the survey.
 - Introduce the interviewer and organization to the client.
 - Begin building trust between the client and interviewer.
 - Familiarize the client with the survey and its purpose.
 - Stress confidentiality, voluntary participation, and that the respondent may ask questions.
 - Gain permission to proceed.
- Do not let anyone else answer for the respondent. Consider ‘who is in the room’ and how another person’s presence could potentially affect the answers given. Whenever possible, try to interview the respondent alone, without anyone else in the room.
 - If the respondent insists that they do not know the answer to a question but that someone else can answer it, speak with that person *after* completing the rest of the survey.
- Ask the exact questions (as written) in the exact order they appear on the survey form.
- For questions with multiple answers to choose from, do NOT read the answer choices. Ask the question, and match the response to the appropriate answer choice. If it is not clear which answer choice to select, ask follow up questions until it is clear which answer to record.
- Probe (*not prompt*) to increase accuracy/clarity & completeness
 - Probing is asking follow up questions to help the interviewee arrive at an answer.
 - Prompting (which we do not want to do) is suggesting answers
- Listen carefully to determine relevant information.
- Show interest. Pause. Repeat the question if necessary. Repeat the reply to stimulate the client to say more, or to recognize an inaccuracy.

- Record answers in boxes/correct places.
- Verify answers when possible by noting the conditions in the home. Be careful not to make false assumptions – just because you do not see a cow, does not mean they do not own one. Ask follow-up questions if an answer is in doubt but do not imply that the client is lying to you.
- Avoid unnecessary reinforcement: “Oh, that’s very good!” or “That’s too bad!” It is your job to remain neutral during the interview.
- If a client is unresponsive, then politely conclude the interview when appropriate and explain the situation to your Field Supervisor.

Ending the Interview

- Thank the client.
- Tell them that they have provided important/helpful information for the study.
- Answer any questions.
- Quickly proofread the completed questionnaire. Find and correct errors. Clarify answers with the client if necessary.
- Report to the Field Supervisor as necessary.

Do’s and Don’ts

Good interviewers will be sure to do the following:

- Be very familiar with the survey questions and answer options
- Project a confident, self-assured manner (to the degree that it is culturally appropriate)
- Try to keep interruptions to a minimum
- Read the questions exactly as written
- Probe when multiple answers are possible or initial responses are unclear
- Record responses completely and properly

Good Interviewers will also be sure not to do the following:

- Get involved in long explanations of the study
- Justify or defend what you are doing
- Suggest an answer or agree or disagree with an answer
- Let someone else answer for the respondent
- Rush the respondent
- Patronize the respondent or dominate the interview
- Make assumptions
- Fill in false answers

Sample Interviewer Training Schedule

This schedule is meant to be illustrative of the activities to be included in interviewer training. It may vary depending on the number of Interviewers you need to train, the number of languages to be used, and your access to members of your target population with whom you will conduct practice / pre-test interviews.

Day 1	Day 2	Day 3	Day 4
--Introductions; agenda and logistics for training week. --Purpose, objectives and key concepts of this research and the tools. --Interviewers' roles and responsibilities. --Reviewing the survey in the first language. --Techniques of survey interviewing.	--Review the surveys in each of the languages and make consensus changes in writing. --Pre-test surveys in the local languages.	--Continue pre-test interviews. --Group discussion to reflect on pre-test interviews and lessons learned.	--Make final changes to each of the tools in each language. --Develop interviewer schedules --If necessary, complete any final practice or training sessions.

Practice Interviews / Pre-testing the Survey

Pre-testing the questionnaire with people who are similar to the sampled respondents is an essential step in finalizing the questionnaire translation and fine tuning the interviewing skills of your team. No matter how experienced your staff may be, the pre-tests almost always bring out potential problems, inconsistencies or other sources of bias and error. Practice interviews should only be conducted with individuals who have NOT been selected for the survey sample, but who are also members of the target population. If interviews are to be conducted in both rural and urban areas, pre-testing should also take place in both rural and urban areas.

Pre-testing a questionnaire in the field should involve everyone on the survey team. The pre-test is essential for finding weak points in translations and errors in the logistical plan, as well as identifying the need for additional field staff training. The quality control persons and data processing team also should process the pre-test questionnaires so that they can have an opportunity to check their systems and procedures to identify problems that may arise.

Pre-testing provides an opportunity for your team to make corrections before doing the actual survey. The pre-test also allows your Project Manager and Field Supervisors to learn more about the time and resources required to locate and interview respondents. Pre-testing is done as part of the interviewer training.

CHAPTER FIVE – FIELD WORK: INTERVIEWING AND QUALITY CONTROL

In order to ensure that field work goes smoothly, it is important to develop standard practices and procedures for both field and home office staff. Advanced preparation can save you a lot of time in the field, where time is precious.

Field Supervision

Adequate field supervision is a key to effective survey implementation. Roles, procedures and responsibilities should be clear to all team members. Consider the following questions:

1. How often should Interviewers check in with their Field Supervisor and/or turn in completed surveys?
2. How will the Interviewers communicate with their Field Supervisor? Will they receive cell phones, phone cards, or airtime for their own cell phones?
3. What should the interviewer do if someone on their list is unavailable or refuses to participate in the survey? Do they have a list of alternates? Should they check in with their Field Supervisor?
4. When should Interviewers take breaks / lunch?
5. How much per diem do Interviewers receive and when do they collect it? Must they turn in receipts?
6. Are there designated meeting places for the team? At what time does the day begin? End?

Time Management

Traveling and locating those to be interviewed can easily take up more time than the interviews themselves. Adequate advance planning can minimize travel time and maximize the chances that the interviewer will be able to easily find the client.

Your field staff that work with the clients in each region should be consulted concerning travel routes, locating the clients' homes, and the best time of day to find the clients. These same staff members also should inform the clients ahead of time about the interview if at all possible to increase the chances that clients will be available at the intended time.

An option that can be time-saving but has other drawbacks is to schedule the interviews to occur immediately after meetings that are already taking place with the client (such as a regular group loan meeting or a training session). This method can reduce greatly the amount of time your Interviewers spend traveling and trying to find those on the sampling list, but will likely prevent the interviewer from being able to observe the housing situation. Note that the interviewer must still be able to speak with the respondent alone and away from the group to maintain confidentiality. A happy medium is to arrive at the end of the weekly meeting, be introduced to the sampled client by the loan officer, and then accompany that client back to their own home for the interview.

Quality Control

The best time to correct mistakes is immediately after they occur. *Thus, the most important step in the quality control process is carried out by the Interviewers and Field Supervisors while they are still out in the field.* Once the survey forms have returned to the central processing location, it will be expensive to try and return to the field to re-interview anyone. The basic steps of the quality control process in the field are:

- Interviewer confirms that the person to be interviewed is on the sampling list before beginning the interview.
- Interviewer uses knowledge of survey and how the questions relate to one another to look for inconsistencies in responses, probing for clarity when necessary.
- Interviewer looks over the survey form immediately after completing the interview, looking for missing, incomplete, or inconsistent entries, or any writing which a data entry staff member might not be able to read. If the interviewer sees that any information is missing, they might be able to remember what the answer was or be able to ask the interviewee before they leave. Since the poverty level calculation is based on all responses to the survey, every question must be completed.
- Field Supervisor meets the Interviewers after each one or two interviews to collect the surveys and look them over to make sure they are complete, easy to read, and have no inconsistencies. If the Field Supervisor sees a number that is not clearly written, the Interviewer should be instructed to clarify it as soon as possible, and reminded to make sure all the information is legible. The supervisor also should check the sampling list to make sure the sampled person was interviewed or that the substitution was done correctly.
- Even though some survey teams may prefer to wait until all of the interviewing is completed before entering any data, it is recommended to enter at least a few surveys each day, even if you delete the data later. Doing this helps you find erroneous responses or patterns that you otherwise would not have caught when looking over the form in the field. Chapter 6 provides instructions on how to enter the data and run frequencies to see the distribution of the answers.

Troubleshooting Scenarios:

- At the end of the day, it becomes apparent that a few survey forms have missing data and it is not possible to find the respondents and ask them for the information. The Interviewers do not remember the responses, either.
 - Solution: Throw the questionnaires away. Since the poverty level calculation is based on the responses to all of the questions, all of the answers must be provided. Find additional names of clients to interview that fit into the sampling plan to make up for the loss (if you do not already have alternates selected in the sampling plan). Try to complete the interviews in the remaining days scheduled for field work, but if you have to spend an extra day to make up the loss, then start planning the logistics for the extra day.

- After all of the interviewing is completed, the data entry staff notice problems with several of the survey forms and are unable to use them. It would be expensive and difficult to return to the field to interview appropriate replacements.
 - Solution: Your oversampling should be able to make up for the loss in order to reach the goal numbers of clients in your sampling plan. If the oversampling does not make up for the loss, then consult the Help Desk for further instructions. This scenario is a good example of how important it is to both oversample and to test entering data in Epi Info or CSPro before the interviewing process is complete.

Document Control and Tracking

It is critical to ensure that there is a system in place for collecting and tracking all documents, including systematizing survey numbers that are written on each questionnaire prior to the interviews, and a using a log book for tracking the status of each document.

Internal Audit

Training Interviewers and other team members well, selecting a representative sample, and properly tracking survey documents and data processing are all essential for accurately assessing the poverty level of your program beneficiaries. The surveys should be reviewed for completeness and consistency at several levels of quality control. However, survey review and document tracking will not necessarily detect or prevent someone from taking shortcuts or “gaming the system.” Lazy Interviewers could interview the wrong person or even spend the afternoon sitting under a tree and filling out the surveys by themselves. Interviewers could also mistreat the interviewees and damage the reputation of your program.

To prevent such problems, an internal audit should be an essential but distinct component of the PAT implementation process. A member of the Core Team, or other organizational staff member with relevant experience should conduct “spot tests” by re-interviewing a random sample of those who have already been interviewed by the implementation team. Ideally, this would include at least 2 interviews of clients that were already interviewed by each interviewer. The “re-interview” should include:

- Confirmation that the person interviewed is on the sample list and was actually interviewed.
- Comparison of the responses with the completed survey from the initial interview.²⁷
- Additional questions to make sure the client was treated properly and that proper interviewing techniques were followed.

In addition, the internal auditor should:

- Review the sampling process to ensure the proper procedure was followed and that the sample selected is truly representative. This can include comparing the demographic information from the sample to the demographic profile of clients already captured in the program MIS database.
- Examine the document tracking system, including a random selection of survey forms in various stages of the process to ensure that each step of the process is tracked, including each level of quality control.

²⁷ Note that you should not assume that discrepancies in responses are necessarily due to negligence on the part of the Interviewer since it is possible that the respondent changed their mind and answered differently for whatever reason. Use follow up questions at the end of the interview to determine the reason for discrepancies.

CHAPTER SIX – DATA PROCESSING - Epi Info and CSPro

This chapter outlines the data processing and poverty level calculation process. After interviewing a representative sample of the institution's clients, the practitioner needs to enter the collected data into a database (the Data Entry Template for your country) and run the program that calculates the percentage of respondents who are 'very poor' for reporting to USAID. The process entails work done by two separate data entry personnel and a Data Processing Coordinator. The directions below describe the process, step-by-step, from quality control at the end of the field work to reporting the results to USAID, ending with a section on troubleshooting.

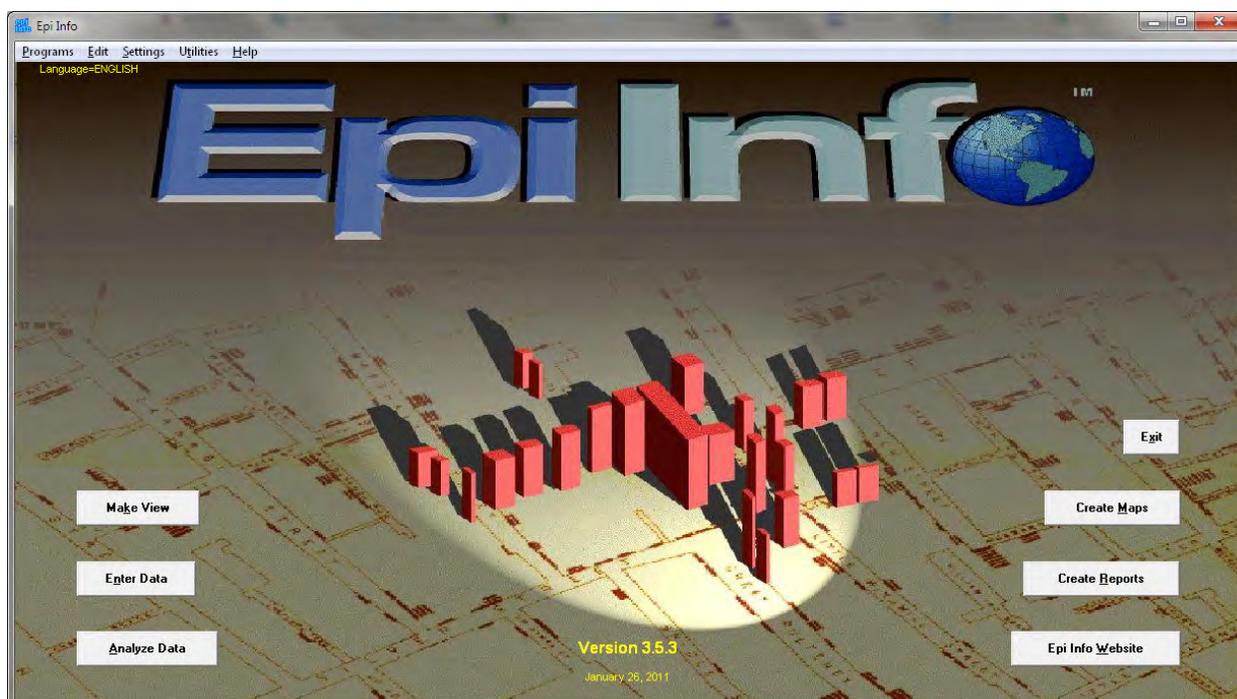
PAT data entry and analysis are done in either the Epi Info or CSPro software packages. PATs were originally designed in Epi Info. In 2010, in response to implementer feedback and the desire to provide increased functionality, IRIS tool developers began to create PATs for use within the CSPro platform. New tools created after July 2010 were developed in CSPro and retrofitting of older PATs began at that time as well. As of December 2010, 8 of 33 PATs use CSPro for data entry and processing. New tool development and conversion of older tools to CSPro will continue in 2011. ***An updated list of which software is used by each PAT, is maintained at www.povertytools.org/tools.html.*** As of December 2010, Epi Info versions of Data Entry Templates that have been retrofitted into CSPro are still available on the website. Consult the PAT Help Desk (pathelp@iris.umd.edu) if you have questions about which Template version you should use to meet USAID reporting requirements.

Quality Control at Main Office

Once enough paper surveys have been completed to meet both the sampling and oversampling requirements for the institution, a final round of quality control should be done prior to data entry. Since the calculation of the poverty level of the respondents will be incomplete unless all of the responses are entered for each survey, the documents should be reviewed one more time to ensure they are complete. Any forms that are missing any answers should be sent back to the Field Supervisors to return to the field to either complete them or interview replacement candidates. If a survey is found to have missing information and it is too late to return to the field, the survey must be discarded. *It is very important that missing values are not replaced with a "0" (or any other placeholder value) if that was not the correct answer. An incorrect "0" value will skew the data and invalidate the results of the poverty calculation (this applies to missing responses only – in some PATs, instructions will tell you to fill in a "0" if the question does not apply to a member of the household, for example).* Supervisors should use the quality control boxes at the top of the survey to help keep track of which forms have been checked for quality.

The final round of quality control is also an excellent opportunity to clarify any handwritten answers that may be difficult for data entry processors to read, and to ensure that the stack of paper surveys is in order by survey number. Keeping the surveys in numerical order will facilitate looking up the correct value in a particular survey later during the analysis process.

Epi Info



In this section you will learn all of the steps needed to process your PAT survey data in the Epi Info database, including:

1. Becoming familiar with the Epi Info software
2. Organizing your Data Processors
3. Entering, cleaning and analyzing the data
4. Reading and interpreting your results

What is Epi Info?

Epi Info is free software created by the Center for Disease Control and the World Health Organization. It is capable of creating forms and databases and performing data entry and analysis. Epi Info can do many things above and beyond what you will need it to do. We will focus only on using the menus and functions needed to complete the steps outlined above.

The features that make Epi Info suitable for use with the USAID PATs include its adaptability, simple point and click procedures, customized data entry with controls to limit errors, cost (free), and the ability to compare data to minimize errors. **Both the software and its user manual can be downloaded at www.cdc.gov/epiinfo.**

System Requirements²⁸

Windows 98, NT 4.0, 2000 Requirements

- A 200-megahertz (MHz) processor (Pentium compatible or greater) is recommended but not required
- 32 MB of Random Access Memory (RAM); 64 MB for Windows NT 4.0 and 2000
- At least 260 megabytes of free hard disk space to install; 130 megabytes after installation

Windows XP and Vista Requirements

- Pentium 300 megahertz (MHz) processor or faster (300 MHz is recommended)
- At least 260 megabytes of free hard disk space to install; 130 megabytes after installation
- For XP users: the XP Service Pack 3
- For Vista users: the “DHTML Editing Control for Applications” must be installed

We recommend that you follow along and complete the steps described in the following pages to familiarize yourself with Epi Info. When we get to the section on comparing databases (under the heading “Compare Data”) you will need to download files (“Using Epi Info – practice templates”) from the training website (www.povertytools.org/PATtraining) or follow along by looking at the screenshots.

Saving Files

Epi Info automatically saves changes you make to any file you are working on. As a result, it is necessary to create copies of any file that you intend on editing BEFORE you open it in Epi Info. We recommend creating a “Backup” or “Saved Epi” file. You will want copies of every file you create along the way to your final file, so it is easy to trace your modifications. Having an appropriate name for any file you have modified is also useful. After you open and change a database file in Epi Info, rename it AS SOON AS YOU ARE FINISHED to reflect the contents of the new file. We will give you examples of how to do this throughout this lesson.

First Steps

1. Download the most recent version of Epi Info from www.cdc.gov/epiinfo.

If you already have Epi Info on your computer, make sure you have the most recent version (version 3.5.3 was released January 26, 2011). The version number appears at the bottom of the screen when you open the program (see the yellow text on the first screen).

2. Download the Survey Tool²⁹ and Data Entry Template for your country from www.povertytools.org/tools.html. Save them in a folder together.

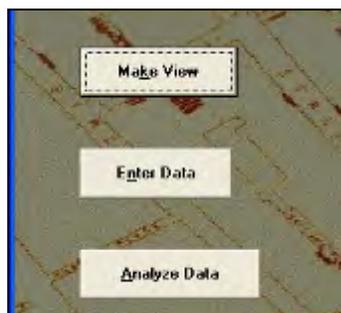
²⁸ Epi Info does not run under other operating systems such as Linux or Mac OS.

²⁹ The PAT survey can also be downloaded from this website but is only used in its printed form for collecting survey data. Data is entered directly into the template in Epi Info, not through the survey which is in Excel.

3. Unzip (decompress) the Data Entry Template and save the extracted file in the same folder holding the Survey Tool and zipped (compressed) Data Entry Template. **DO NOT open the extracted file yet! It uses the same extension (.mdb) as a Microsoft Office Access application though the file is not designed to work with Access. Opening it may cause your computer to stall or stop working entirely. You will open it once you are in the Epi Info program.**
4. *Optional:* you may delete the zipped Data Entry Template folder. From this point forward, we will refer to the file you unzipped as your “Template.”
5. *Recommended:* make copies of the Survey Tool and Template and save them in a separate folder. This will save you the trouble of having to download them again if you make changes and need to see the original documents.
6. Open Epi Info by double-clicking on the icon on your desktop or selecting the program from the Start menu.

Basic Functions

We are going to focus on the “Make View,” “Enter Data,” and “Analyze Data” functions on the left side of the screen.



Make View is used to create and edit survey forms or Data Entry Templates.

Enter Data is where your data entry occurs. Each survey is saved as a separate record. Here, you can search for specific surveys or answers within a survey in addition to basic data entry.

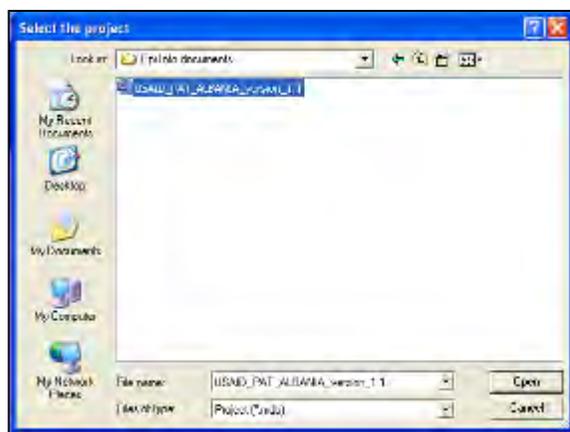
Analyze Data is used to run the poverty calculation program that is built into the Data Entry Template. You can also use it to perform quality control checks, merge databases created by different Data Processors, and conduct additional analysis on survey data.

Make View

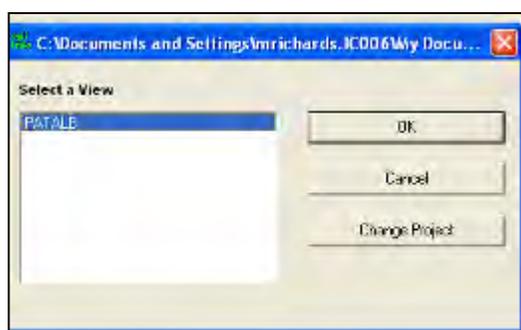
This function is used to create survey forms and Data Entry Templates. We have already created your Data Entry Template for you. However you may want to translate it into a local language if it will help your Data Processors. Follow these steps:

1. Click “Make View.” A new screen will open.
2. Click “File” then “Open.” Another screen will open.

3. Find the folder where your “Survey Tool” and “Template” are stored (see page 52, “First Steps”). Select your Template and click “Open.”



4. Next you will be asked to “Select a View.” There should only one option (*PATALB* in this example). Select it and click “OK.”



Your survey should open. It will look slightly different than the Excel version of the survey you have already looked at.

5. To edit any of the text (for translation, for example), double-click the text you want to change. A new window will open up and you can change the text in the “Question or Prompt” window. When you are finished, click “OK.”

You should NOT change any other information in this window. Doing so may cause the data analysis program to not work. However, note the “Field Name” section. Later, when you want to look at survey results or find specific records, you will have to choose which Field Name(s) to use. “Survnum” is an important one to remember because you will be sorting and merging databases and records by their survey number.

Also note that many surveys are on more than one page. Use the navigation in the top-left hand corner to switch between pages.

Any changes you make are saved automatically. When you are finished, click “File,” then “Exit” at the top of the screen.

Enter Data

This function creates a data table for the data being entered, controls the data entry process by limiting values that can be entered in a specific field, and includes a search function so that specific records can be located.

Your Data Processors will use this function to enter all of your survey results and also to look up records and make edits when you find odd or suspicious values (more on that later). Just like in “Make View,” Epi Info automatically saves your data as you type. To enter survey data, follow these steps:

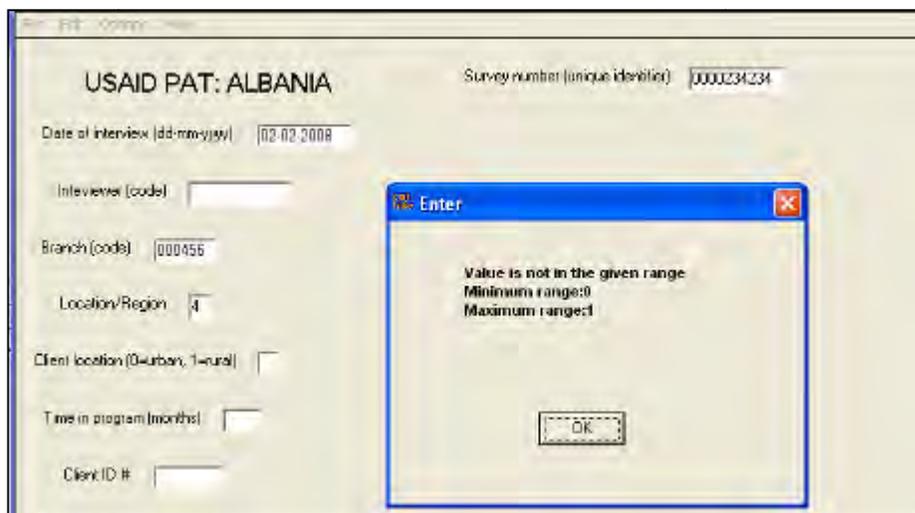
1. Click “Enter Data.” A new screen will open.
2. Click “File” then “Open.” Another screen will open.
3. Find the folder where your Survey Tool and Template are stored. There should only be one file for you to open. Select the Template and click “Open.”
4. Next you will be asked to “Select a Table.” There should only one option (*PATALB* in this example). Select it and click “OK.”

Your survey should open. It will look just like it did in “Make View” without the grid lines. Also, the data entry fields will be white instead of grey.

The screenshot shows a software window titled "Enter" with a menu bar containing "File" and "Edit". The main area is titled "USAID PAT: ALBANIA" and contains several data entry fields: "Survey number (unique identifier)", "Date of interview (dd/mm/yyyy)", "Interviews (code)", "Branch (code)", "Location/Village", "Client location (District, Village)", "Time in region (months)", and "Client ID #". Below these fields is a section titled "Household information: please enter values from the data entry blocks below the roster" with a field for "2. Number of people in the household". A "Data" button is located at the bottom left of the form area.

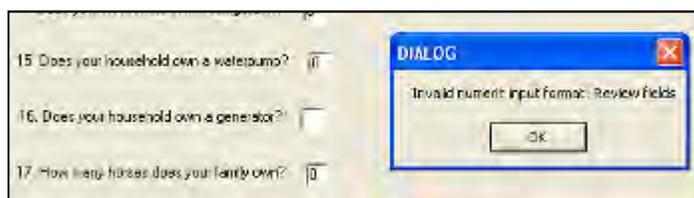
To enter data, click the first empty field (“Survey Number” in this example) and start typing. To move to the next field, use the “Tab” or “Enter” key.

When we created this Template, we included ranges that will limit responses to reasonable answers. If a Data Processor tries to enter a number outside the acceptable range, they will get an error message. In the picture to the right, I tried to enter “2” as the “Client location.”



6. Use the “Tab” or “Enter” key to complete the data entry. When you get to the bottom of a page and press “Tab” or “Enter”, you are immediately brought to the next page. You may also use the page navigation in the top-left corner of the screen. When you finish answering the last question, pressing “Tab” or “Enter” brings you to the next (new) record.

7. If any data fields are left blank, you will get an error message when you get to the end of the survey. You must then GO BACK to the survey you just completed and fill in the missing data. You CANNOT leave any fields blank³⁰.



8. When you are finished entering all of your data, close the window. Remember, your changes are saved automatically.

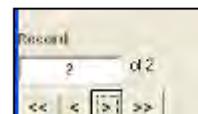
9. Recommended. Copy and re-name saved databases to reflect who completed the data entry and which records were included. For example, if Ruth entered records 1 through 67, she could name the saved file “USAID_PAT_RUTH_1_67.mdb.”

Later, you can make further changes, like “USAID_PAT_RUTH_1_107.mdb” if you’ve merged the databases with records 1-67 and 68-107 together or “USAID_PATdata_all.mdb.” Be sure to save earlier copies of the databases in case you need to go back to them or something happens to your most recent copy.

³⁰ In the next section, we will explain how to deal with missing or illegible data on survey forms during data entry.

Other Things You Need to Know About “Enter Data”

The numbers in the bottom-left corner of the screen tell you which record you are currently looking at (in the order they were entered, NOT with respect to survey numbers) as well as the total number of records in the open database. You can move to different records using the arrows or by entering the record number you want to see in the white field and pressing “Enter.” The double arrows (<< and >>) bring you to the first and last records. Clicking “>>” twice allows you to enter a new record.



There are also several command buttons on the left side of the screen.

Next advances you to the next page in the current record.

Save data can be used to save an incomplete record (by-passing the error message). We do not recommend using this as there is no way of later identifying records that have been saved with missing data. Incomplete records are NOT included in the poverty calculation. Excluding data in this way affects your sampling plan and accordingly, the accuracy of your overall results.



Mark record as deleted is Epi’s way of ‘deleting’ a record without removing it from the database entirely. When a record is “marked as deleted”, it is not included in the final poverty calculation or any other analysis function. The benefit is that you can fix a record “marked as deleted” (after you’ve found missing data, etc) and then “undelete” it.

Find allows you to search for a particular value (such as “survey number”). You will use this function to ‘clean’ your data.

Data Compare

Before we analyze our data, we have to go through a few more quality checks. The first one involves comparing data. We recommend that each survey be entered by TWO different people (Data Processors) into two separate databases in order to reduce error. While individuals can make errors easily (finger slips on the keyboard, enters “222” instead of “22”, etc) it is unlikely that two people will make the same error. Epi Info includes a “Data Compare” function that allows you to compare two databases with the same records.

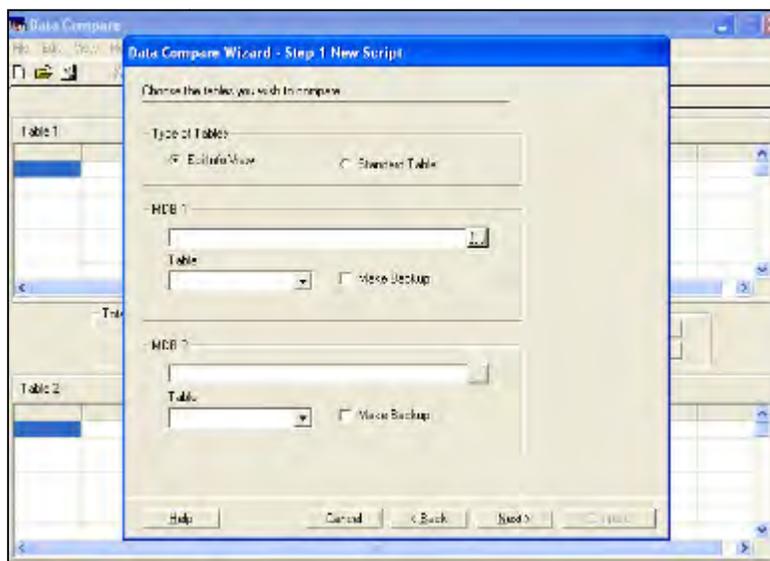
You can download the files named below from the training website and follow along. They are in a folder called “Using Epi Info – practice files” on the course homepage.

After downloading the files (which will be compressed/ziped), unzip each file but do NOT try to open the resulting .mdb files yet.

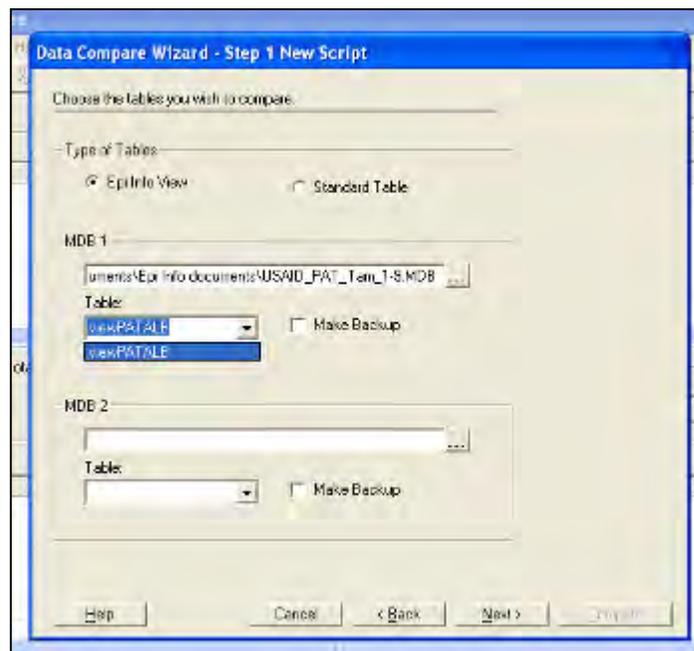
1. The same records (surveys) are entered twice using “Enter Data”, preferably by two different Data Processors, and saved to two separate database files.
2. These databases should be given similar names, to make it easier to compare them. For example, data entry person Tam would save his database as “USAID_PAT_Tam_1-9.mdb”, and Yao would save his as “USAID_PAT_Yao_1-9.mdb.” (1-9 indicates that they both entered survey numbers 1 through 9 into their databases.)
3. These two databases will now be compared to identify differences in survey values entered by Tam and Yao to find errors. To do so:
 - a. From the main Epi Info screen, select the “Utilities” menu near the top left corner of the screen. Select “Data Compare.”



- b. In the “Data Compare” screen, select “File” then “New Script.” The “Data Compare Wizard” will appear on the screen.



- c. Select “Epi Info View” as the “Type of Table.” Then, under “MDB 1,” select the first database (“USAID_PAT_Tam_1-9.mdb”) using the “...” button and click “OK.” Under “Table,” select the view “viewPATALB” (this should be your only option). Do the same thing for “MDB 2,” choosing the database you are going to compare to (“USAID_PAT_Yao_1-9.mdb” in this example), and then the view under “Table.” Click “Next” at the bottom.



- d. A unique identifier³¹ is needed to match records between the two databases for comparison. For our surveys, that identifier is the variable (or Field Name, as we

³¹ A unique identifier is a value that is going to be different in every record but the same across databases. Each record is going to have a different “Survnum” (survey number) but Tam and Yao should have used the same survey numbers in each of their databases since they entered the same data.

described previously) “Survnum.” Check the box next to that variable and click “Next.”



- e. In the next screen (“Step 4 – Fields to Compare”, make sure that all fields are checked (since we want to compare the answers to every question). If they are not all checked, click “Check All” to include all fields. Click “Next.”
 - f. Click “Next” again. (We do not need an HTML report of the comparison, so leave the box unchecked.)
 - g. Click the “Compare” button on the lower right corner to see the results.
4. If the two databases are exactly the same, the screen will say “No Differences” in big letters. You can then re-name one of the databases you just compared to show that the records have been checked and are accurate. For example, “USAID_PAT_FINAL_1-9.mdb.”
 5. If they are not the same, records that have differences will appear under the “Differences” tab. Any record that was entered in one database but not the other will appear in the “Unmatched Records” tab.
 - a. Select the “Differences” tab.
 - b. Table 1 shows records from the first database you entered. In our example, “USAID_PAT_Tam_1-9.mdb.” Below that you can see the data from the second database, “USAID_PAT_Yao_1-9.mdb”
 - c. Values which differ between the two databases/tables are highlighted in yellow. Items being currently compared are in blue.
 6. To correct mistaken values:
 - a. First, make sure you are in “Edit” mode and not “Read-Only” mode. Look above the “Differences” tab. You will see a button that looks like either an Eye or a Pencil.

The Eye means “Read Only.” It changes to a Pencil, meaning “Edit” when you click it once.

- b. If it is clear that a value in Table 1 is wrong, but that the value in Table 2 is right, replace the incorrect value with the value from Table 2. To do so, click on either value and then click on the “Accept Value Table 2” button.
- c. You can also change a value by double-clicking it and typing in the new value.
- d. If there is any doubt as to what the correct value is, go back to the paper surveys and check.**

On the next page, we will walk through the differences between Tam and Yao’s databases.

First, note that we can see the Eye in the tool bar, meaning we cannot make any edits yet. Click it until it changes to a pencil.

Survnum 200801: Tam has entered that there are 15 rooms [in the household] and Yao has entered 1. Looking at the other survey results (6, 2, and 3 rooms), Yao’s answer is probably right so we will click “Accept Table 2 Value.” If there is any doubt, we should go back to paper survey number 200801 and check the answer.

When we accept the correct value (or manually change one to make them the same), the new values will be in green.

Survnum 200805: Looking at *Timeinprog*, which is more likely: that a person has spent 21 or 221 months in the program? Let’s accept Table 1 Value. We could also double click “221”, erase it and enter “21.”

Survnum 200808: Again, do we think this person has spent 43 or 4 months in the program? Let’s say the program has been around for a long time so that someone might have been in it for 43 months. Since we’re not sure which answer is right, we would go back to the original paper survey and find the answer. In this case, let’s say it is 4 and Accept Table 2 value.

There’s another error in this survey. Is the household size (Hhsize) 3 or 33? Again, unless we’re 100% sure, it is best to go back to the paper survey and check. In this case, we will accept 3 as the answer.

Not done yet...at the bottom of the window there is a scroll bar. We have to slide the table to the right to see the rest of the data. There are actually 2 more errors.

The screenshot shows the 'Data Compare' window with two tables side-by-side. The top table, 'Table 1', has the following data:

Survnum	Rooms	Cell	Cai	Phone	Boiler	Fridge	Ypr	Computer	Horsecdy	Generator	Numprimcom	Waterpump	agrieland
200801	1	0	0	0	0	0	0	0	0	1	12	0	5
200805	6	1	0	0	0	1	1	1	0	1	0	1	1365
200808	2	1	1	0	1	0	0	1	0	1	1	1	12
200809	3	0	1	0	0	0	1	1	1	1	15	0	12

The bottom table, 'Table 2', has the following data:

Survnum	Rooms	Cell	Cai	Phone	Boiler	Fridge	Ypr	Computer	Horsecdy	Generator	Numprimcom	Waterpump	agrieland
200801	1	0	0	0	0	0	0	0	0	1	1	0	5
200805	6	1	0	0	0	1	1	1	0	1	0	1	1365
200808	2	1	1	0	1	0	0	1	0	1	1	1	12
200809	3	0	1	0	0	0	1	1	1	1	5	0	12

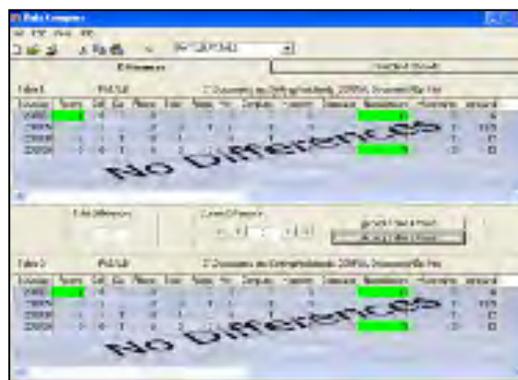
(See how the Rooms values have turned green? That is because we fixed them.)

Survnum 200801: Next, look at the column that says “Numprimcom.” This is the field name of the question that asks for the number of family members who have completed primary school (not including the head of household). 12 or 1? Let’s look at the number of people in the household and see if that gives us a clue. Scroll back to the left (or look at the screenshot on the previous page) and see that there are 12 household members total (11 plus the household head). This tells us that there cannot be 12 members (not including the head of household) who have finished primary school because there are only 11 non-head-of-household household members. So should we accept 12? Well, maybe the interviewer read the question wrong and included the head of household in the total number of people who have completed primary education. We have to go back to the paper survey to find out that yes, the interviewer did make a mistake. We counted up the number of people who have completed primary school and there are 11 NOT including the household head. This means that BOTH data sets had a mistake! Let’s change both values to 11.

Survnum200809: As with the last survey, let’s compare the number of family members who have completed primary school (not including the head of household) with the number of members in the household. There are only 12 people in the household total so we can reasonably assume that 15 is the wrong answer and accept 5 (Accept Table 2 Value) as the right answer.

After we accept this last correct value, we get a confirmation that we're done!

Remember, Epi Info automatically saved the changes you made to *both* databases that you were comparing. Do not forget to rename one of them as a 'final' version, for example, "USAID_PAT_FINAL_1-9.mdb."



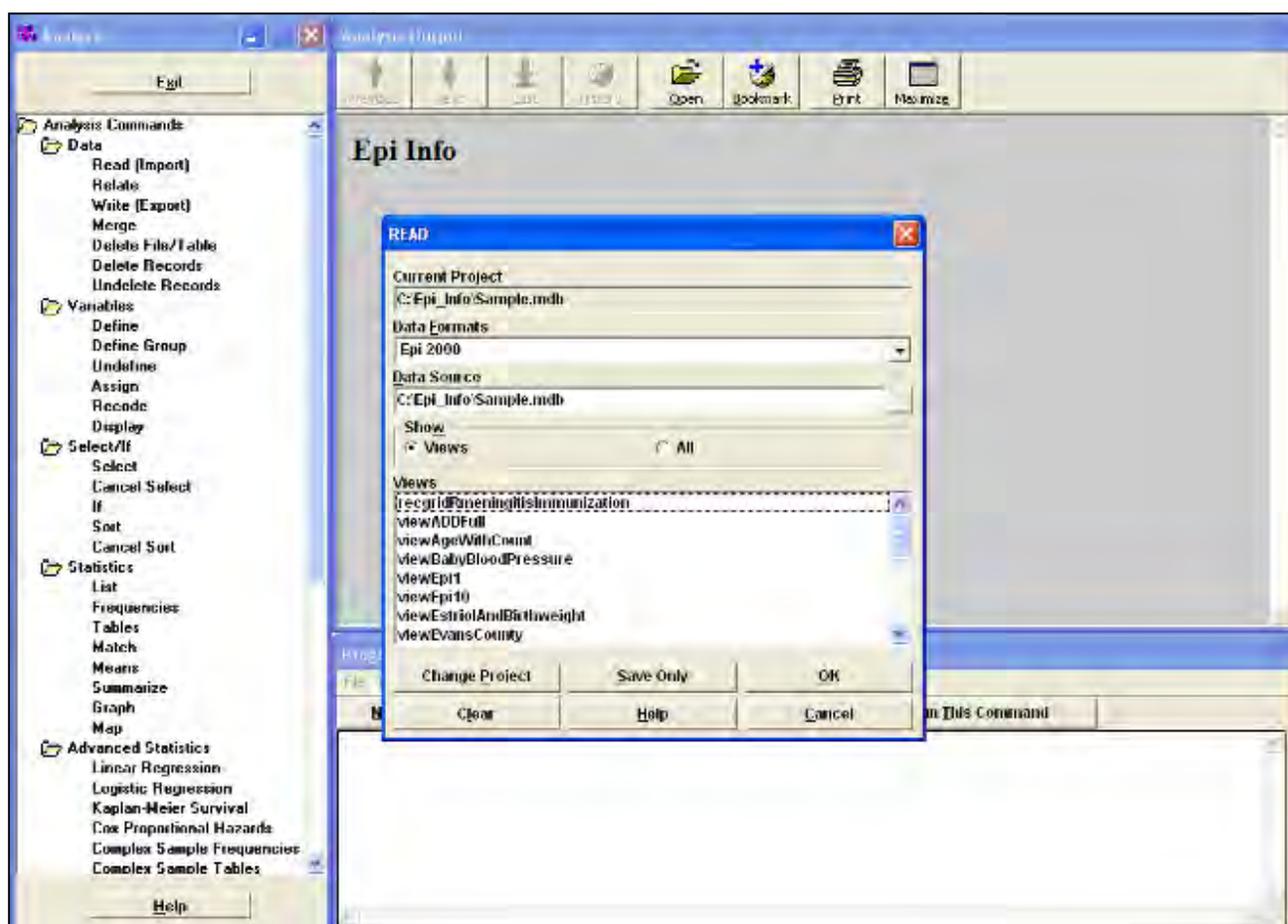
Analyze Data

We're not quite done with combining and checking our data yet but to do the next step, we have to learn how to get into the Analyze Data function so that we can:

- **Merge** two databases with different surveys in them
- Run a **Frequencies** calculation to look for odd or outlying data points
- Run the **data analysis program** to calculate our poverty rate

These are the first steps you will follow to do all of the above:

1. Open the Analysis program by clicking on the “Analyze Data” button on the main Epi screen.
2. Select “Read” from the list of “Analysis Commands” in the left-hand column.



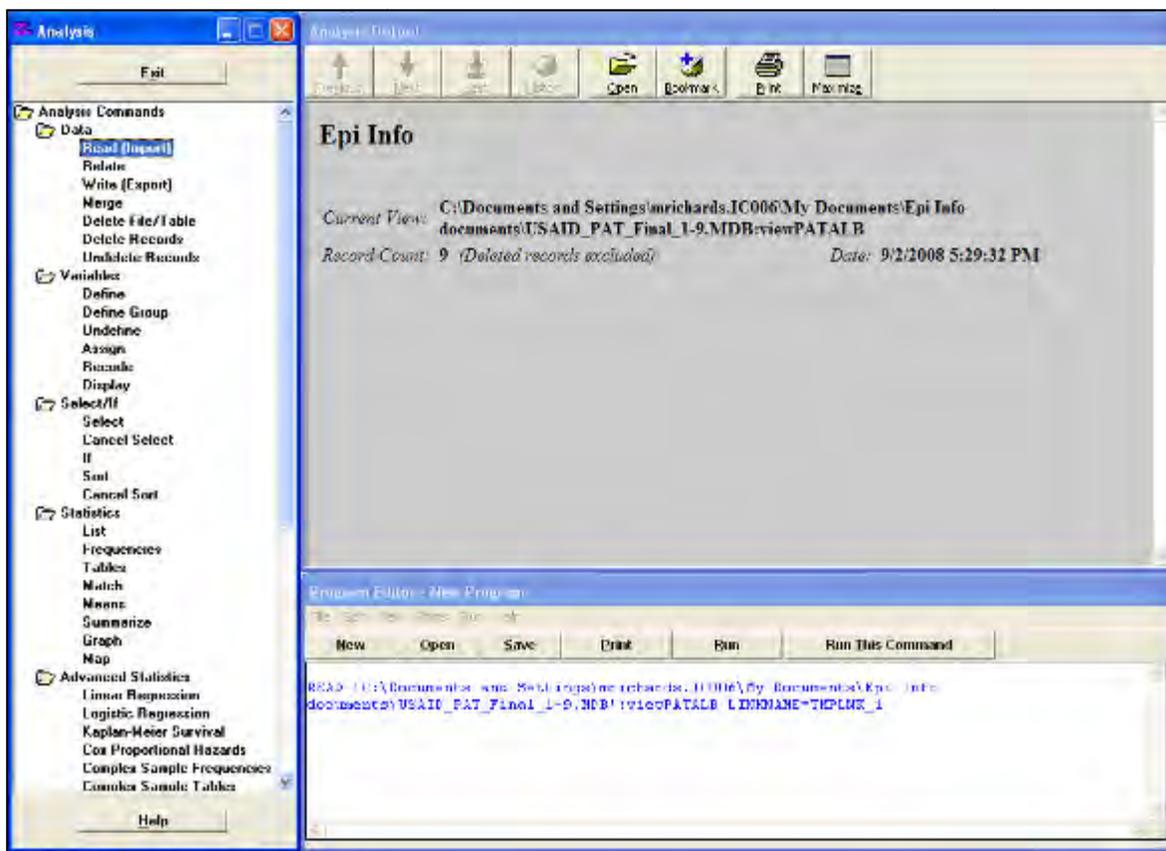
3. Click the “Change the Project” button (bottom-left corner) and find the database you want to use.
4. Click “Open.”
5. Under “Views”, choose the view (it should have the same name as the View you were using previously.).
6. Click “OK.” If you are asked about creating a Temporary Link, just click “OK.”

Merging Data Sets

If different surveys are entered into two different databases, they will need to be merged into one. The “Merge” command is used to perform this function. As an example, we’re going to use “USAID_PAT_FINAL_1-9.mdb” (the file we created after comparing Tam and Yao’s databases) and “USAID_PAT_FINAL_10-13.mdb” (another database containing surveys numbered 200810 to 200813). Merging the two databases means copying the contents of one database and adding them to the other. Be sure you have made backup copies of the files you are going to merge.

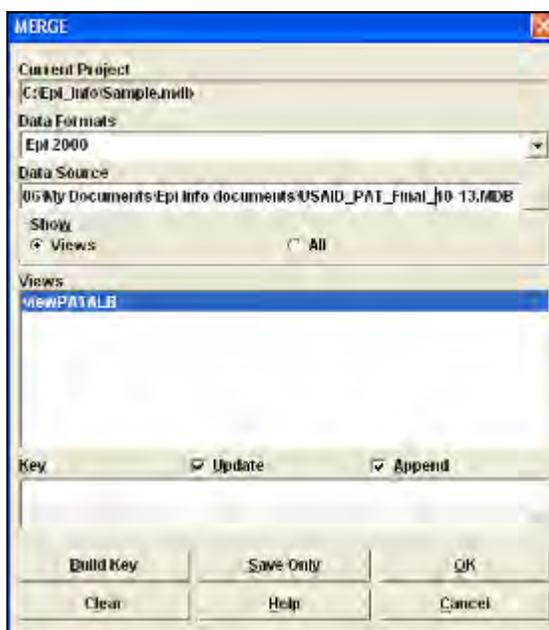
First, make sure that the databases you’re merging have different names (ours do). Put them in the same folder so they are easy to get to.

1. Follow the steps on the previous page to read “USAID_PAT_FINAL_1-9.mdb” into the analysis program. The next screen will look like this. Make sure the “Record Count” is accurate. The records from our second database will be *added* to this database.

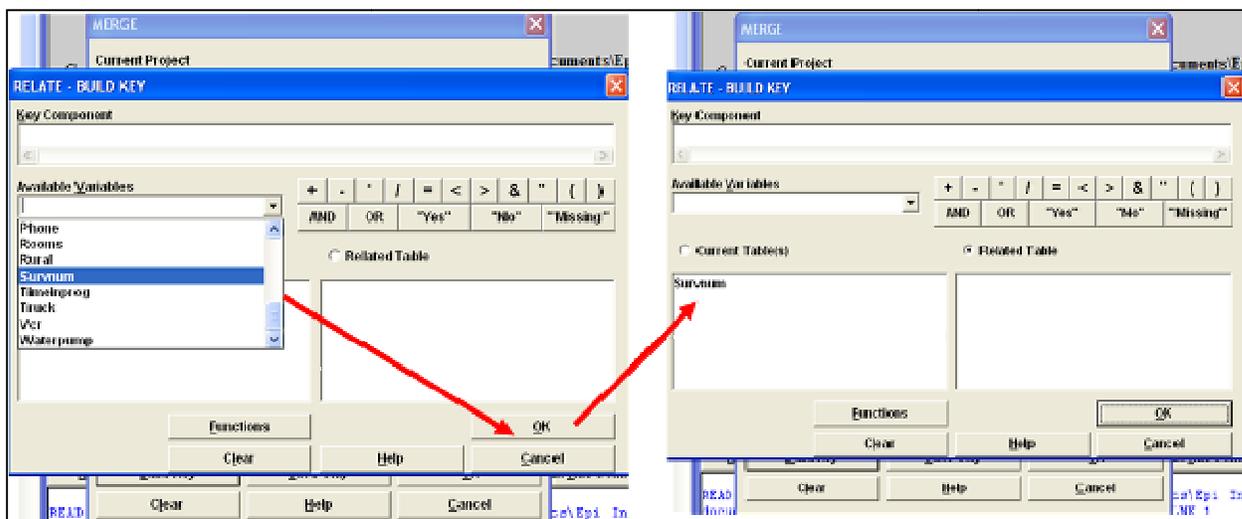


2. From the list of “Analysis Commands” on the left, under “Data”, select “Merge.”
3. Under “Data Source”, click the “Find File” button on the right-hand side (it looks like an empty square), and locate the second .MDB file that will be merged into the first file (“USAID_PAT_FINAL_10-13.mdb”)

4. Click “Open.”
5. Click on the “View” name and confirm that “Update” and “Append” are both checked. “Update” means that if you have any overlapping records (surveys with the same survey number) in your 2 databases, the values from the *second* database will overwrite the values in the first database. “Append” adds new records from the second database to the first database. This is one of the reasons it is important to keep track of all of your files and know which ones contain the most recent and accurate data.

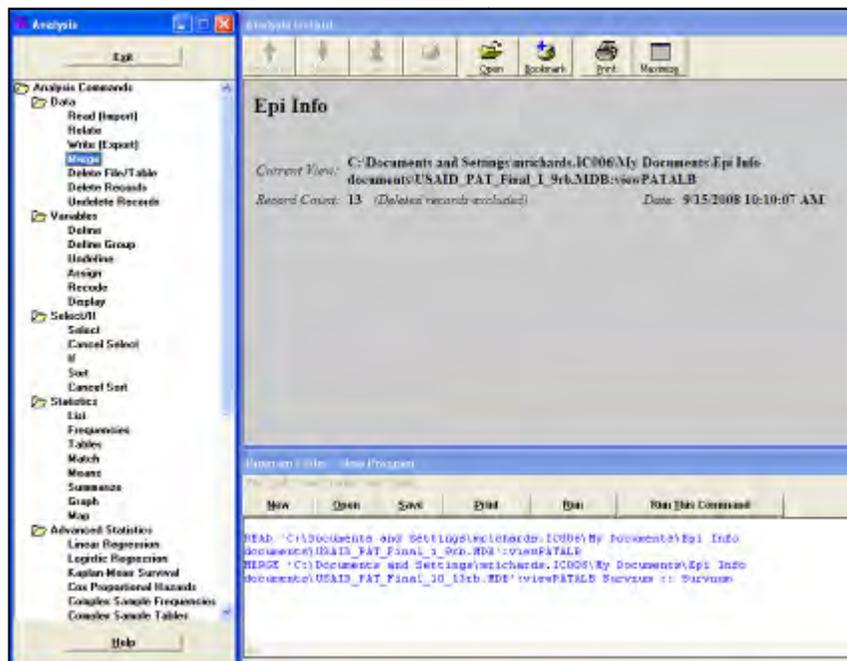


7. Click on “Build Key” in the bottom-left corner above “Clear.”
8. Choose the “Survnum” field from the “Available Variables” drop down list. (Click on the small black arrow on the right of the “Available Variables” field, then select “Survnum” from the list.)
9. Click “OK” for the selection to appear under the “Current Tables” section.



10. Repeat the process by choosing “Survnum” again from the “Available Variables” drop down list.
11. Click “OK.” This time it will appear under the “Related Table” section.
12. Click “OK” again to go back to the “Merge” dialog box.
13. Click “OK” to merge the files.
14. If you are asked about creating a Temporary Link, just click OK.

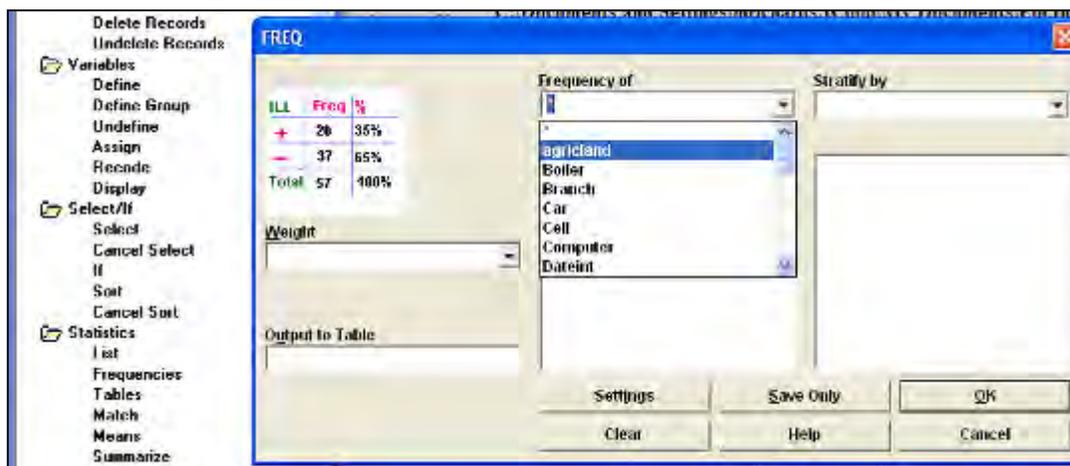
You should now see that the “Record Count” has increased to include the total number of records from both databases. In this case, 4 records from “USAID_PAT_FINAL_10-13.mdb” were added to the 9 records in “USAID_PAT_FINAL_1-9.mdb” to create 13 total records. Remember that now, “USAID_PAT_FINAL_1-9.mdb” contains all 13 records so you will need to rename it something like “USAID_PAT_merge_1-13.mdb.” Notice that I did not include the word FINAL in this file name because we are not done yet. We have one more quality control step to take before we analyze our data.



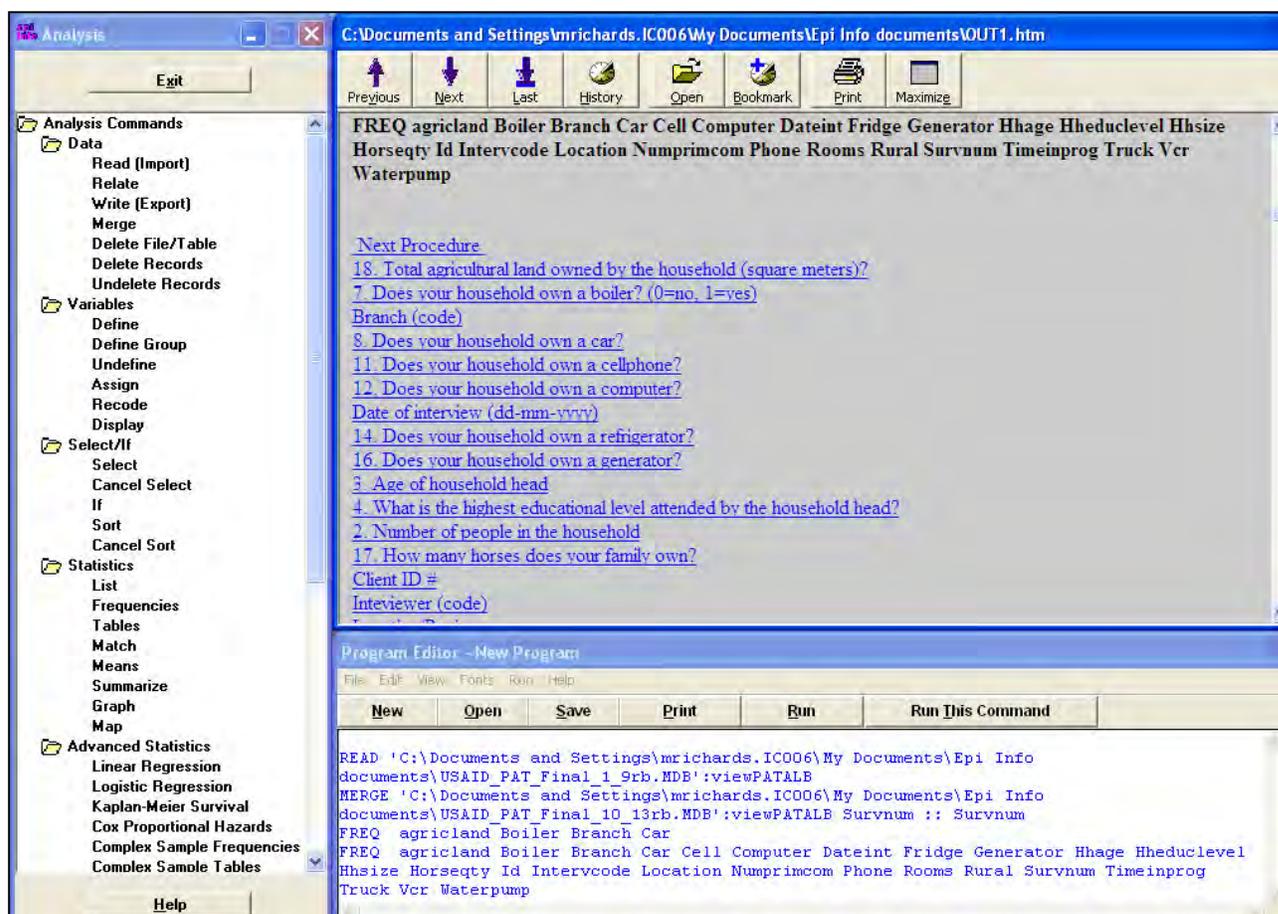
Frequencies (looking for odd / outlying values)

We already talked about comparing two databases with the same records to look for mistakes. In the example we used, we also found a case in which both of the Data Processors made a mistake on the same question on the same survey. Survey errors made by Interviewers (or bad handwriting!) can sometimes be found even after surveys have gone through a number of quality control checks. For this reason, we run “frequency” calculations on our databases to look for survey answers that just do not look right.

1. Follow the steps outlined earlier to get into the “Analyze Data” program and select the database you want to check.
2. Under the “Statistics” heading of the menu on the left, click “Frequencies.” The “FREQ” dialogue box will open.
3. Under “Frequency of”, click the down arrow on the right to see the dropdown list that contains all of the variable/ field names for every question in our survey. You can pick individual codes to look at one at a time or you can select any or all of them. Each time, click the down arrow to view the dropdown list, select a variable name and its name will appear in the white box below. Selecting the asterisk (star *) adds all of the variable names.



4. When you’re done, click “OK.” Your screen will look like this:



The top line (“FREQ agricland Boiler Branch...”) shows you the names of all the fields whose answer frequencies you are looking at. Below that in blue are the lists of the questions associated with each Field Name. They are sorted alphabetically by Field Name. Below that in the Program Editor box, you can see the commands (code) that Epi Info used to generate your results.

To see the results of your Frequencies operation, you can click on the name of an individual question to jump down and see the results of that particular question or you can scroll down to see all of the results. Recall that for yes/no questions (e.g., Does your household own a car?), 0 = no and 1 = yes.

5. Scroll down to Age of household head (Question 2). Do you see a suspicious value? It is possible that a client stated that their household was headed by someone who is 2 years old but we should check. However, since we do not know which survey contained that value, we will have to go back and view the database in *Enter Data* to search for the 2 year old head of household.

3. Age of household head	Frequency	Percent	Cum Percent
2	1	7.7%	7.7%
22	1	7.7%	15.4%
23	1	7.7%	23.1%
32	2	15.4%	38.5%
45	3	23.1%	61.5%
51	2	15.4%	76.9%
54	1	7.7%	84.6%
55	1	7.7%	92.3%
61	1	7.7%	100.0%
Total	13	100.0%	100.0%

Finding specific survey data

1. Go into the “Enter Data” program from the main Epi Info screen. As before, open the database and then the View/Table that contains the record you are looking for.
2. Click on the “Find” button on the left side. The “Find Record” screen will be displayed.

Find Record

Choose search field(s)

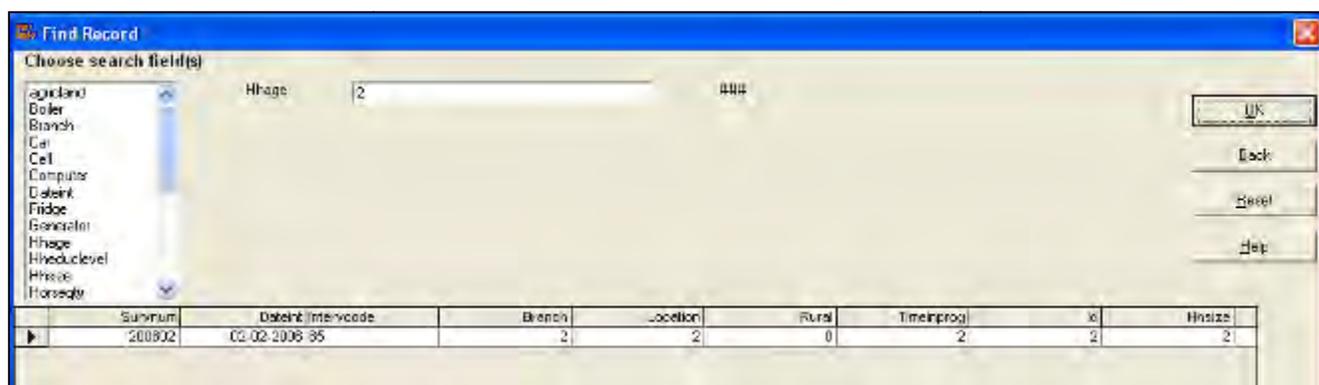
Branch
 Car
 Cell
 Computer
 Dateint
 Fridge
 Generato
 Hhage
 Hhdulevel
 Hhsize
 Hhsectr

Find:

###

OK
Back
Reset
Help

3. Under “Choose search field(s)”, select the Field Name you need to look for (Hhage, in this case).
4. Enter the value you are looking for (“2” in this case).
5. Click “OK.”



6. You will then see a list of all the records with a Household Head Age of 2. In this case, there is only one. Double click the arrow to the left of the record you want to look at and you will be brought to that record.
7. When you find odd values, you will have to decide if they can be fixed or not. As we've done before, go back to the paper survey to see if the correct data is there. If not, you can check with the Field Supervisor and / or Interviewer to see if they can make sense of the mistake or if it is possible to go back to the client and clarify their answer. If none of these are possible, you must delete the survey. **DO NOT use a value of zero for missing answers.** Missing data ALWAYS requires that the survey in question be thrown out unless the answer can be verified and added. Remember you can always use the "Mark as deleted" option to remove a record from being included in poverty and other calculations. This is a good option if you cannot correct your data right away and will have to come back to it because you can always "Undelete" using the same button.
8. For the purpose of our practice exercise, let's not make any changes (leave "2" as the household age for survey 200802).
9. After you've corrected all the mistakes you find, be sure to re-name your file so that you know that all the data has been fixed ("USAID_PAT_FINALmerge_1-13.mdb").

Once we've corrected our surveys a final time and deleted any records that cannot be fixed or used for some other reason, we are ready to run the analysis program!

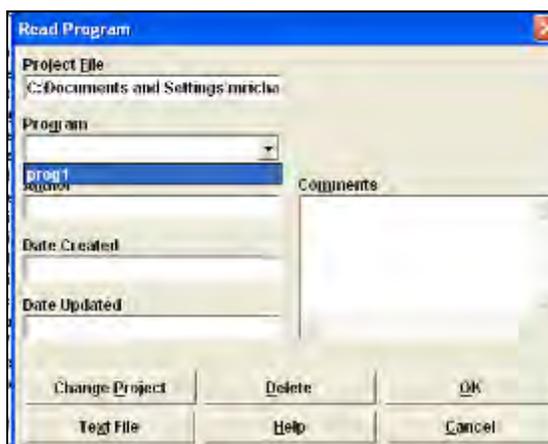
Calculating the Poverty Level

Before running the poverty calculation program, make sure you have...

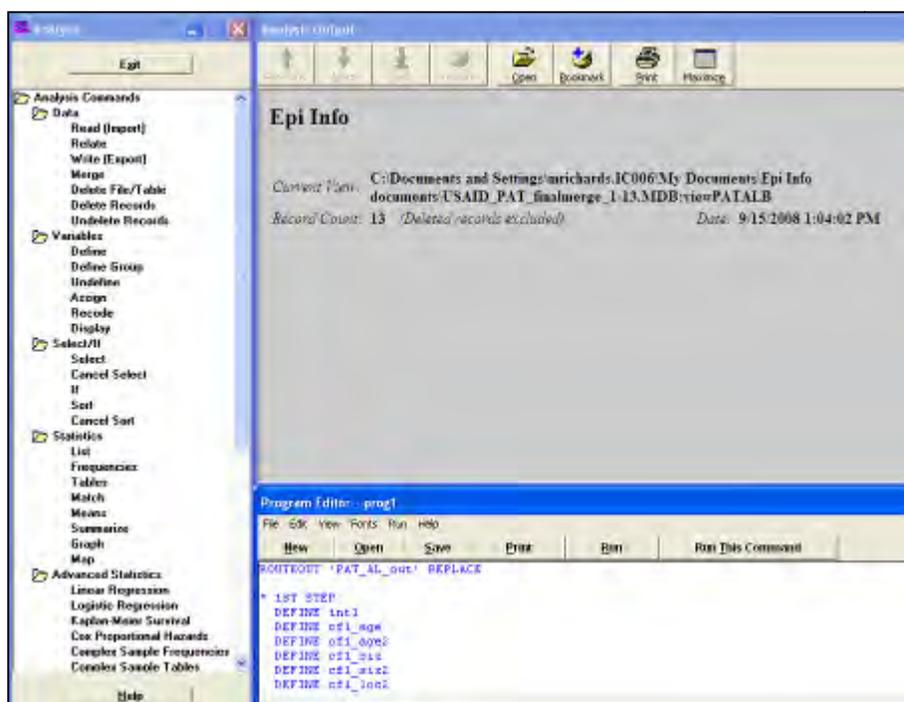
- Entered all survey into the database using “Enter Data.”
- Compared and fixed errors in databases for which you have ‘double entered’ data from the same surveys using the “Compare Data” function.
- Merged databases containing different surveys so that you have ONE database containing ALL of your survey data.
- Run frequency checks on all survey data. This can be done before or after merging databases.

The process for running the poverty calculation program is simple.

1. Follow the steps outlined earlier to get into the “Analyze Data” program to read and select the database you want to check.
2. Make sure the number of records displayed matches the total number of surveys that were completed in your sample.
3. Look at the “Program Editor” near the bottom of the screen. Click the “Open” button to open the “Read Program” dialogue box.
4. Verify that the correct “Project File” is selected. If not, select “Change Program” to find your database and open the file.
5. Back in “Read Program”, click the arrow next to the white box under “Program.” You will have one option: prog1. Select that and click “OK.”



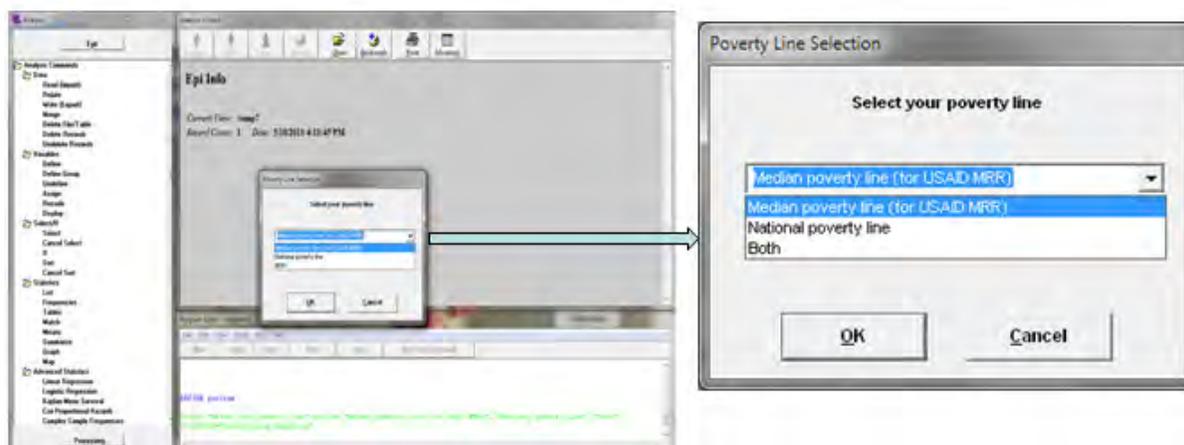
6. You will now see that many lines of commands have been added to the “Program Editor” section. **DO NOT change anything in that area** or it will prevent you from successfully calculating the poverty level of your client sample.

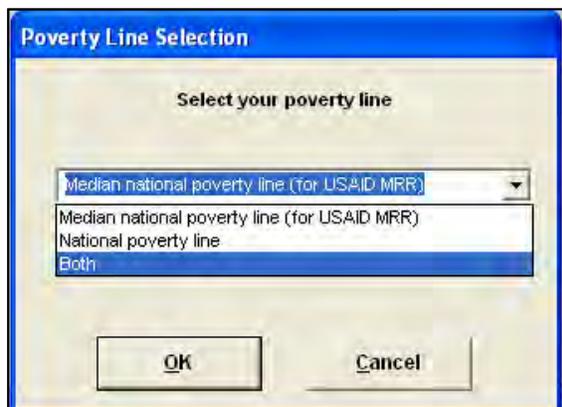


7. In the “Program Editor” section at the bottom of the screen, click “Run.” This will execute all of the lines of the program.

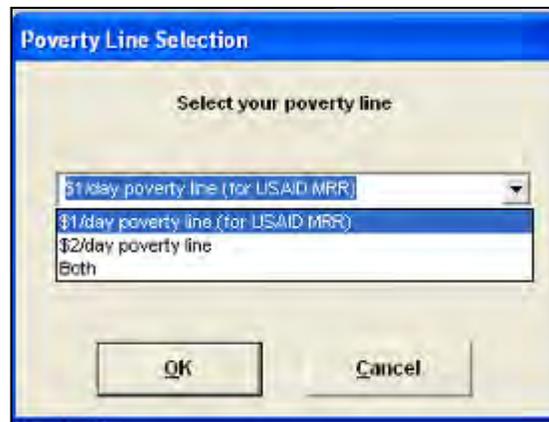
The newest PAT Templates will ask you to select whether you would like to see your country’s poverty line, its extreme poverty line, or both. In the example below, the poverty lines used are the Median and National poverty lines. Remember that some PATs will use the Median national poverty line and the National poverty line and others will use the \$1/day line and \$2/day line instead. At the current time, the Albania PAT does NOT give you this choice so if you are following along, you will not see the boxes below, you will just see your results after you press RUN.

An example of what will happen with the newer PAT templates after RUN is pressed:





PAT using the Median/ National line



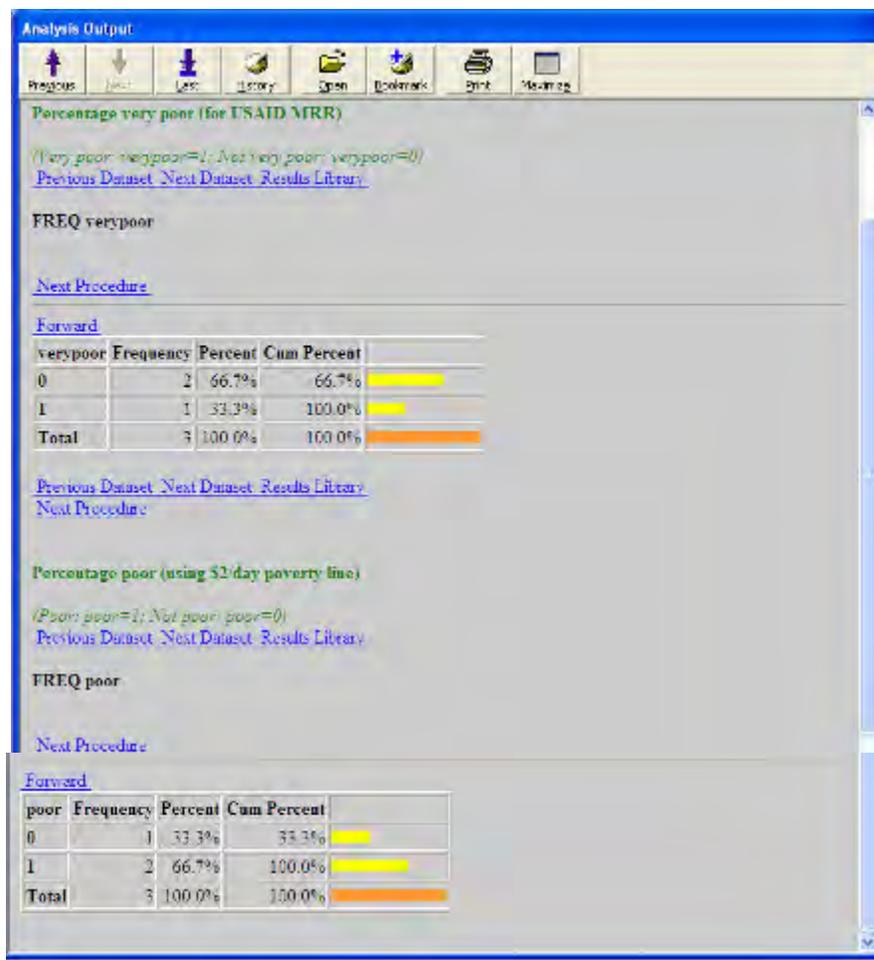
PAT using the \$1/day , \$2/day poverty line

8. Your results will look like this. The table of values shows you the number and percent of your clients who are “not very poor” (“0”) and “very poor” (“1”). In this example, we see from the second row in the table that 12 households (92.3%) were determined to be “not very poor.” One household (7.7%) is “very poor.”

verypoor	Frequency	Percent	Cum Percent
0	12	92.3%	92.3%
1	1	7.7%	100.0%
Total	13	100.0%	100.0%

95% Conf Limits
0.640% 99.8%
1.02% 36.0%

9. We've changed the formatting of the results in Epi Info slightly for the newest tools. They will look more like the image on the right. You will have to use your scroll bar to see all the results as they may not fit on one screen. At the top, we see the results of the "very poor" calculation using the \$1/ day line (1 person out of 3 – 33.3% - is **very poor**). Below that, we see the results of the \$2/day calculation (2 people – 66.6% - are **poor**).



As we mentioned before, Epi Info has a number of functions that you may find useful in analyzing your survey data but that are not essential for fulfilling the reporting requirement. Look at some of the other functions in the left-side menu of the Analysis function.

Troubleshooting

What if I need to translate the Data Entry Template?

If you have translated the survey questions and would like the data entry screen to match what is on the survey used in the field, you can change the questions (field labels) that appear on the Epi screen to the translated version. Use the Make View function as instructed on page 54.

How do I email my database files?

Sometimes email software (such as Microsoft Outlook) does not allow a database file created in Epi Info (.mdb) to be opened as an attachment. To get around this, and also to greatly reduce the size of your attachment, be sure to use WinZip or another file compression software package to compress the database. Then it can be attached to an email and opened by the recipient. The database files will reduce in size significantly using WinZip (sometimes to as little as 10% of the original file size). This makes it very easy to transfer the database files by email or other means.

To make the PAT database smaller:

1. Go to the main Epi Info screen.
2. On the toolbar, click “Utilities” then “Compact.” The **Compact** screen will open.
3. Under “MDB file to compact,” click on the open file .
4. Find the folder where your database is stored.
5. Click on your database file, then click “Open.”
6. Under “Name of compacted MDB file,” the name should be automatically filled in for you. *Do not change the name of the file.*
7. Leave the “Password” blank.
8. Click “OK.”
9. There should be a window that will pop out saying “Compact complete.”
10. The compacted file should be in the folder where your original database is stored.

How do I analyze my data using other software?

In order to carry out the poverty calculation program built into the PAT databases, the data must be entered and processed using Epi Info (or CSPro, if that is the format used for your country’s Data Entry Template).

However, if you wish to do additional analysis on the data using another statistical software package, you can export the data into Excel, and then import it into another software program, such as SPSS.

To export your PAT data into another format:

1. Open the Analysis program by clicking on the Analyze Data button on the main Epi Info screen.
2. Select READ from the list of Analysis Commands in the left-hand column.
3. Click the Change Project button and find the database file (.mdb) that you would like to export. Click Open.
4. Under Views, choose the desired view. Click OK.

5. Select Write (Export) from the list of Analysis Commands in the left-hand column.
6. In the WRITE inset box, make sure the 'All' box is checked above the list of Variables, and then select the Output Format into which you will export the data. (Click on the small black arrow to the right of the field to see the drop down menu of options). Excel 4.0 is a good option for then importing the data into SPSS, but there are also options to export the data from Epi Info into Access, dBase and Paradox.
7. After selecting the Output Format, type in the name of the File Name you would like to use for saving the new file.
8. Click OK, and your new file will be saved automatically in the C:\Epi Info folder. Then you can open your preferred statistical software and import the Excel data file for further analysis (or do additional analysis in Excel).

CSPro

The list of PATs that use the CSPro software can be found at www.povertytools.org/tools.html.

This section will explain all of the steps needed to process your PAT survey data using the CSPro software, including:

1. Becoming familiar with the CSPro software
2. Entering, cleaning and analyzing the data
3. Reading and interpreting your results
4. Exporting your data

What is CSPro?

The **Census and Survey Processing System (CSPro)** is a free software package for entering, editing, tabulating, and disseminating data from censuses and other surveys. It was developed by the United States Census Bureau, Macro International, and Serpro, S.A .

The features that make CSPro suitable for use with the USAID PATs include its ability to mimic the design of a paper survey, simple point and click procedures, customized data entry with controls to minimize errors, and the ability to compare data to minimize errors.

CSPro is capable of doing many things above and beyond what you will need it to do to meet USAID reporting requirements. This tutorial focuses on using the modules and menus needed to process your PAT survey data. Both the software and its user manual can be downloaded at www.census.gov/ipc/www/cspro/.

System Requirements

- Pentium processor
- 256 MB of memory
- 70 MB free disk space
- Windows 98SE, ME, NT 4.0, 2000, XP, Vista or 7³²

Installation³³

1. Go to www.census.gov/ipc/www/cspro and click the link to download CSPro.
2. You will be prompted to complete a form with basic information about yourself. If you do not wish to receive email updates from CSPro developers, be sure to uncheck the related box just above the “Submit Form” button.
3. After you submit the form, you will be given several options for installing the software. Whichever option you select, you will need to download and run the installation file (ending in “.exe”).

³² CSPro does not run under other operating systems such as Linux or Mac OS.

³³ Add reference for CSPro Manual

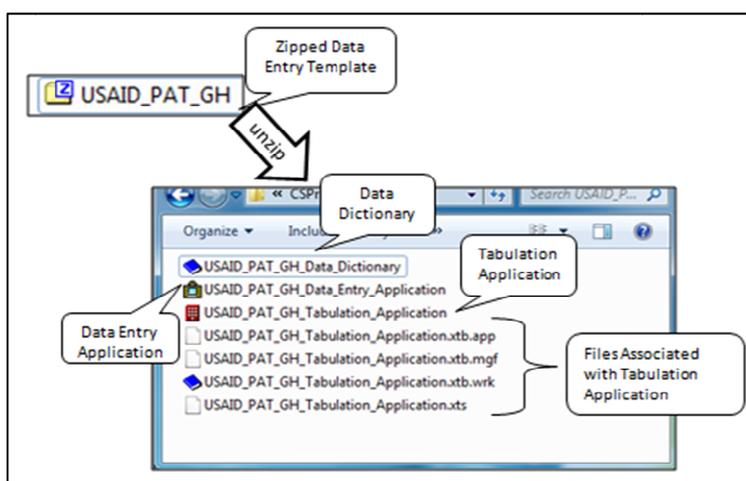
4. After you have downloaded the appropriate file(s), open the .exe file and follow the instructions to install the software.
5. *Selecting language for installation:* During installation CSPro allows you to choose the language that will be used for some parts of the system. *Menu options, dialog boxes, and the help system are only available in English.* After installation, if you want to change the language setting you must rerun the installation program.
6. *Selecting components for installation:* During installation you will see several options for the components of CSPro you would like to install. Choose **CSPro (all components)**. After installation, if you want to change the components installed, you must rerun the installation program.

Downloading the Data Entry Template

1. Download the Data Entry Template for your country from www.povertytools.org/tools.html.³⁴
2. Unzip (decompress) the downloaded Data Entry Template and save the extracted files. *Optional:* you may delete the zipped Data Entry Template folder that you downloaded.
3. *Recommended:* Make copies of the Data Entry Template (and Survey) and save them in a separate folder. This will save you the trouble of having to download them again if you make changes and need the original files.
4. *Recommended:* Download and unzip the CSPro practice files available at www.povertytools.org/CSPro.html. You can use these files to follow along with the examples in this manual.

CSPro Overview

This image shows the contents of the unzipped Data Entry Template folder.



³⁴ The PAT survey can also be downloaded from this website but is only used in its printed form for collecting survey data. Data is entered directly into the data entry template in CSPro, not through the survey which is in Excel.

This tutorial will focus on the following CSPro components: “Data Entry Application”, “Data Dictionary” and “Tabulation Application.”

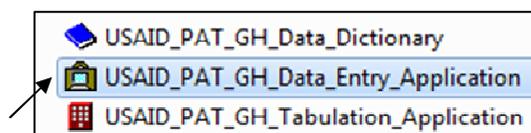
- The “**Data Entry Application**” is where data entry occurs. Each survey is saved as a separate “case”. Here, you can search for specific surveys or answers within a survey in addition to performing basic data entry. You can also modify previously entered cases.
- The “**Data Dictionary**” is used to compare the contents of two data files to identify the differences, concatenate (join end-to-end or merge) two or more data files, tabulate frequencies to check for outliers, and export data to files that can be imported into spreadsheets or databases.
- The **Tabulation Application** module is used to summarize the results of the poverty analysis program and provide an indicative summary of the relationship between poverty and household characteristics. The files associated with the tabulation application are not to be explicitly opened or deleted; they are necessary for the Tabulation Application to run. Though you will not directly need to access these files, here are descriptions of their uses:
 - **.xtb.app**: The logic file contains the logic statements that control the tabulation application.
 - **.xtb.mgf**: The message file stores message text and an associated message number.
 - **.xtb.wrk**: This is the secondary dictionary. It contains variables and data items created for tables which are not part of the main data dictionary.
 - **.xts**: This specification file contains tables, dictionary items and value sets and other information which defines a set of tables.

The screen shots and examples that follow use the practice data entry template. *However you can use any country’s Data Entry Template to follow along with the tutorial.* Practice files with which you can follow the examples on your own computer can be downloaded at www.povertytools.org/CSPro.html.

Data Entry Application Overview

The Data Entry Application is the electronic format of the paper survey. To access the data entry application follow these steps:

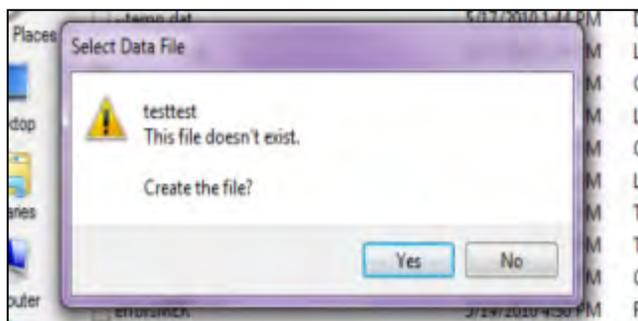
1. Double click on the CSPro Data Entry Application file in the folder in which you unzipped your data entry template (“USAID_PAT_GH_Data_Entry_Application” below). It is the file that has a computer icon next to it.



2. The **Select Data File** window will open. To create a new file for data entry (called an “input data file”), navigate to the folder in which you want to store the data entry file

(preferably the same folder that contains the unzipped data entry template) and type in a name for the data entry file. Then click **Open**.³⁵

- It is recommended that the input data file is named to reflect who completed data entry and which survey numbers were included. For example, if Ruth is entering survey numbers 1 through 67, she could name the file “USAID_PAT_RUTH_1_67.” You can re-name the data entry file later but first read the section “Renaming the Input Data File.”
- When you are asked if you want to “Create the file?” click **Yes**.



The data entry screen will open.

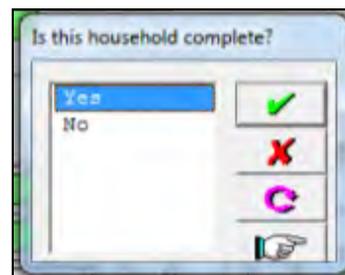
³⁵ If you want to add cases or make edits to an existing file, navigate to that file and click “Open.”

- The **File Tree** shows all the surveys (by survey number) that have been entered in the data file.
- The **Toolbar** is used to access the four modes of operation within Data Entry.
 - Add – Entering new cases³⁶. 
 - Modify – Modifying previously entered cases. 
 - Pause – Temporarily stop adding or modifying cases. 
 - Stop – Stop adding or modifying cases. In this mode the data entry form is not visible. 
- Fields are the areas of the data entry form into which data is entered. The color of each field changes based on the status of data entry.
 - **White**- The cursor or mouse has not moved onto that field during data entry.
 - **Green** - The cursor or mouse has passed through that field for data entry input. All fields in a survey must be green in order for a survey to be included in the poverty calculation.
 - **Yellow** - The cursor or mouse has passed through that field for data entry input and you have moved backwards over the field.
 - **Gray** - The field is protected and data entry cannot occur for this field.

Entering Data

To begin adding data click  (add) on the toolbar (or select **Modify** from the **Mode** menu).

1. Begin by entering the survey number – the cursor automatically appears in the first empty field. To move to the next field, press the **Enter** or **Tab** key. Users have experienced problems with using the mouse to move from field to field so it is recommended that you only use the **Enter** or **Tab** keys. Note that CSPro automatically saves the data entered into the data entry form.
2. After entering basic client information in the top section, you will enter the Household Roster information in the bottom half of the screen. In Column A, the cursor moves from top to bottom. In Columns B-E (all columns but A), the cursor moves from left to right.
3. Once all household member names have been entered, press the **Enter** or **Tab** key twice. The **Is this household complete?** window will open to confirm that the names of all household members have been entered. If you click **No**, CSPro will return to the first blank field in Column A. If you click **Yes**, CSPro will automatically move to the first field in Column B. You will then enter the data for each individual, one by one (left to right).



³⁶ A case corresponds to a survey (questionnaire). Each case must have a distinct survey number.

4. **IMPORTANT:** In Column A you must enter in the name or other identification of each member listed on the survey. You will not be able to enter in data in Columns B-E for household members not entered into Column A.
5. After you have entered data into the last applicable field of Column E, press the **Enter** or **Tab** key. CSPro automatically moves to the next page of the data entry form. To navigate between pages within the data entry form, use the **Page Up** and **Page Down** keys (not the arrows).
6. After you have entered data into the last field of the data entry form, pressing the **Enter** or **Tab** key automatically opens a new data entry form for the next case. You can now begin data entry for the next survey.
7. If you have to verify a response and cannot continue with data entry, you can save an unfinished case. To save a partially entered survey you must at minimum enter the survey number. To stop data entry for this survey, press **Ctrl+S**; from the **Mode** menu, select **Stop**; or click  (stop) on the toolbar. The **Stop Adding** window will open; click **Partial Save**.



In the **File Tree**, partially added cases are marked with a  next to the survey number. If you save a partially created case, you must reactivate the **Add mode** session to continue with data entry.

8. When data entry is complete, click  (stop) on the toolbar or select **Stop** from the **Mode** menu. The changes to your input data file are automatically saved.

Prevention of Missing Values and Consistency Checks

The template intentionally does not allow blank fields and limits the values that can be entered into some fields in order to minimize errors.

For example, the Ghana survey requires the interviewer to designate the household as being in one of ten named regions. Therefore if, as in the example, you were to enter a value above 11 into the “Region” field, you would get an error telling you the value is “Out of Range”.

You cannot bypass the error message by clicking off of the field. If you try to bypass the error message, you will get an error saying “Value is out of range or missing. Value must be between x and y . Verify Response.” Until an in-range value is entered, you cannot move forward with data entry. You also cannot move forward if the field is left blank. You have two options to proceed if you have a blank or incorrect answer. You can stop the survey and return to it at a later date or

you can enter “98” for “no answer given.” If you choose the former, follow Step 7 on page 85. If you choose the latter, see below.

Entering 98 or “No Answer Given”

If you enter “98” for “no answer given” in any of the fields, you will receive a warning once you are at the end of the survey that will prompt you to contact the supervisor and remind you that this survey will not be included in the poverty calculation. To exit the warning, press F8. This case will be saved as a “partial save.” Once you have collected the missing information, you will need to open the case in the “Modify Mode.” You will be asked if you would like the survey opened to the last position. **IMPORTANT:** *Do not select “Yes,” as the program will freeze. Select “No,” then tab to the field you need to change.*

Changing Incorrectly Entered Values

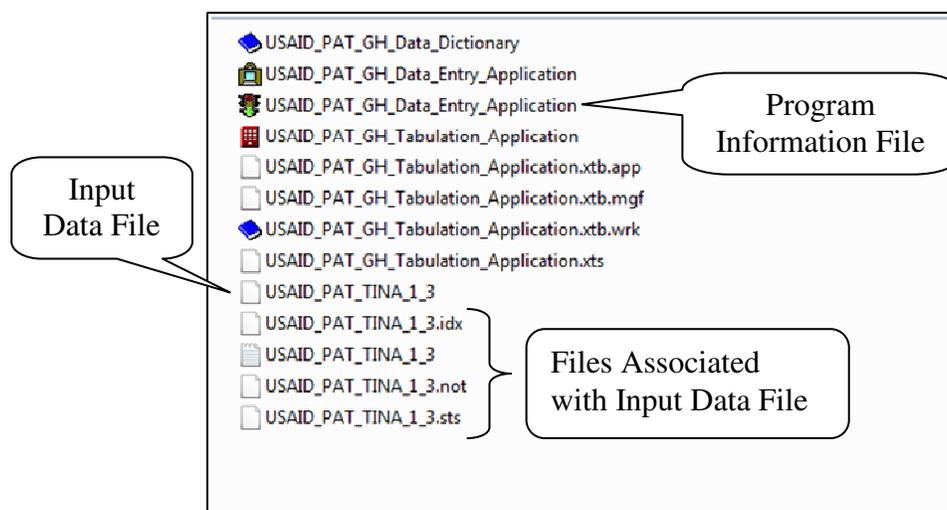
To change an incorrectly entered value during data entry, follow these steps:

1. When you realize you must go back and change a previous answer, first complete entering data in the field you are in but *do not press the **Enter** or **Tab** key.*
2. Using the mouse, click to the field with the incorrect value. After correcting the value, press the **Enter** or **Tab** key until you reach the next blank field. Continue with data entry.

Accessing Data That Has Been Saved

When you want to access a saved input data file (to add new cases or make changes), you must first open the Data Entry Application (the one with the computer icon), then open your input data file in the dialogue box that opens.

When you first create a new input data file, four associated files are automatically created and saved in the same folder. The four associated files have the same name as the input data file but different file type extensions. **IMPORTANT:** Do not move, rename or delete the four associated files. If at any point you change the name of the input data file, you must also change the names of these four associated files to match the name of the new input data file.



Note that there is a file (a text file) with the same name as the input data file – neither has a file type extension attached to it. The icon for the input data file you will access looks like a sheet of plain paper. The text file looks like a sheet of ruled (lined) paper. You will not need to directly open this or any other the associated files but you can find descriptions of their purposes below.

- **.idx:** The data file index stores the case Ids and their physical location in the data file.
- **.sts:** The data file status file stores information about which cases are partially entered and the last field entered.
- **.not:** The notes file stores all the notes entered by data entry operators for the corresponding data file.
- **text file:** The text file is a .log file. This file stores operator statistics generated during data entry.

Program Information File: The Program Information File, also noted in the previous image, acts as a shortcut to open your last-used input data file. It stores the name of the application or tool, the data file to be used and any runtime parameters specific to the application or tool. Always double check which data file you have opened when accessing the data entry template using the Program Information File.

Modifying Data and Adding New Cases

1. Open the Data Entry Application.
2. Select and open the Input Data File containing the data you wish to modify.
3. Double click on the survey number you wish to modify OR click the survey number once and from the “Mode” menu, select “Modify.”
4. The “Survey Number” field will be automatically selected. Press “Tab” or “Enter” to advance to the field you wish to modify. All fields before the field you wish to modify should be green. DO NOT mouse click on the field – you must use the “Tab” or “Enter” key to move from field to field. While you may be able to click and modify, doing so will cause the change to not be reflected in the poverty prediction.
5. Make the edit and then press the “Enter” or “Tab” key **until reaching the end of the survey**. This step is crucial to ensuring that changes are recorded properly. To add a new case to an existing Input Data File, open the file and from the “Mode” menu, select “Add Case.”

Renaming the Input Data File

1. Highlight the Input Data File and its four associated files.

USAID_PAT_MARY_100_102	6/9/2010 1:27 AM	File	1 KB
USAID_PAT_MARY_100_102.idx	6/9/2010 1:27 AM	IDX File	3 KB
USAID_PAT_MARY_100_102	6/9/2010 5:09 PM	Text Document	1 KB
USAID_PAT_MARY_100_102.not	6/9/2010 1:27 AM	NOT File	0 KB
USAID_PAT_MARY_100_102.sts	6/9/2010 1:27 AM	STS File	1 KB
USAID_PAT_TINA_100_102	6/9/2010 1:37 AM	File	1 KB
USAID_PAT_TINA_100_102.idx	6/9/2010 1:37 AM	IDX File	3 KB

2. Right click on the Input Data File. When the next window appears, click **Rename**. The Input Data File should now be highlighted.

USAID_PAT_MARY_100_102	6/9/2010 1:27 AM	File	1 KB
USAID_PAT_MARY_100_102.idx	6/9/2010 1:27 AM	IDX File	3 KB
USAID_PAT_MARY_100_102	6/9/2010 5:09 PM	Text Document	1 KB
USAID_PAT_MARY_100_102.not	6/9/2010 1:27 AM	NOT File	0 KB
USAID_PAT_MARY_100_102.sts	6/9/2010 1:27 AM	STS File	1 KB

3. Type in new file name. Then press the Enter key. The name of the input data file and the four associated files will change to reflect the renaming.

Compare Data Tool

Use the practice files located on www.povertytools.org/CSPro.html to follow along with the next sections. As you will be modifying data by doing so, you should first make a copy of the entire folder.

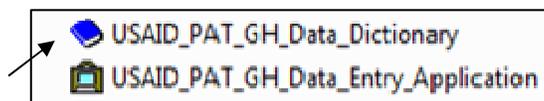
Before analyzing your data, you should go through a few quality checks. The first one involves comparing data. In order to reduce data entry error, it is recommended that each survey be entered by TWO different data processors. While individuals can make errors easily (e.g., finger slips on the keyboard, entering “222” instead of “22”) it is unlikely that two people will make the same error. CSPro includes a **Compare Data** tool that allows you to compare the contents of two data files that contain the same cases (identified by Survey Number) and locate the differences.

The same cases (surveys) are entered twice using the Data Entry Application, preferably by two different data processors, and saved to two separate data files.

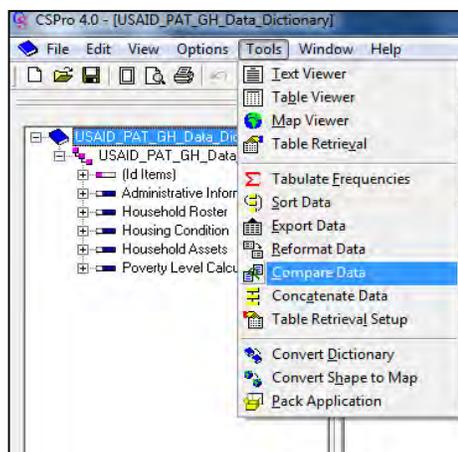
For example, data entry person Mary might save her data file as “USAID_PAT_GH_Mary_100_102”, and Tina would save hers as “USAID_PAT_GH_Tina_100_102” (100-102 indicates that they both entered surveys numbered 100 through 102 into their data file.)

To identify differences between Tina and Mary’s surveys, follow these steps.

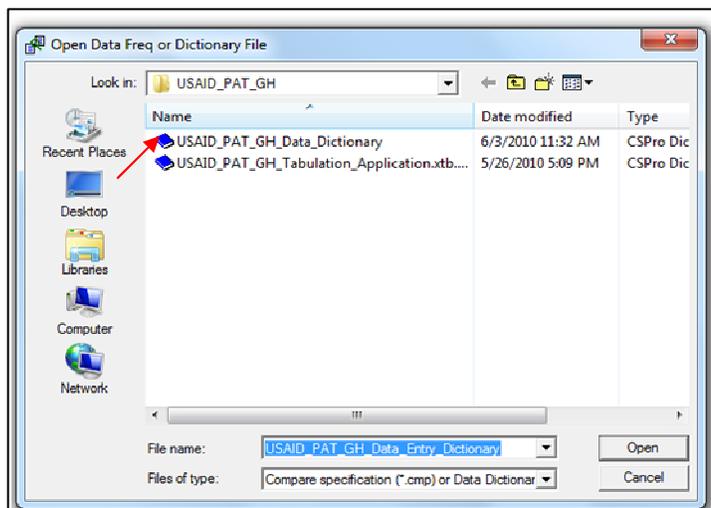
1. Open the Data Dictionary by double-clicking on the CSPro Data Entry Dictionary. It is the file with the book icon  next to it in your data folder.



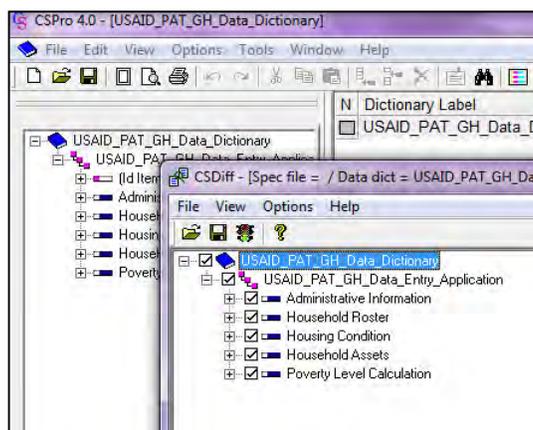
2. Select the **Tools** menu near the top of the screen, then **Compare Data**.



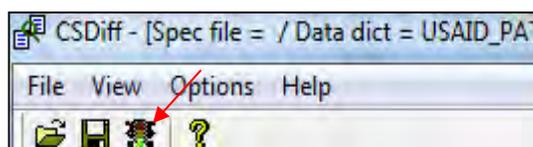
3. In the file menu, select the dictionary file and click **Open**.



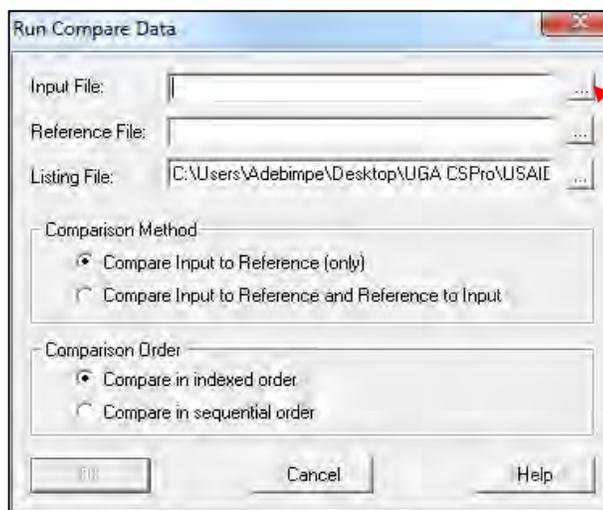
4. In the **CSDiff** window, select the variables you wish to compare. If you want to compare all variables (likely), click on the box next to the dictionary icon to select all. If you want to select specific questions within a category, click on the “+” symbol next to the tick box and the questions within the category will appear.



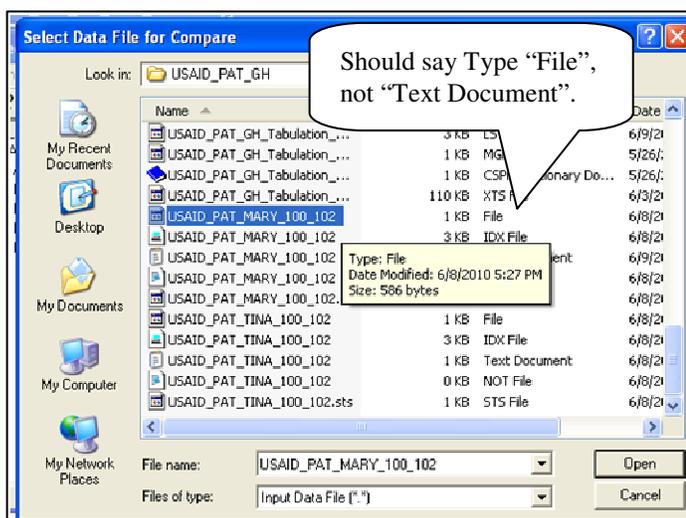
5. Click on the stoplight icon  in the toolbar. The **Run Compare Data** window will appear on the screen.



6. Click on the grey box to the right of **Input File:** and a new window will open.

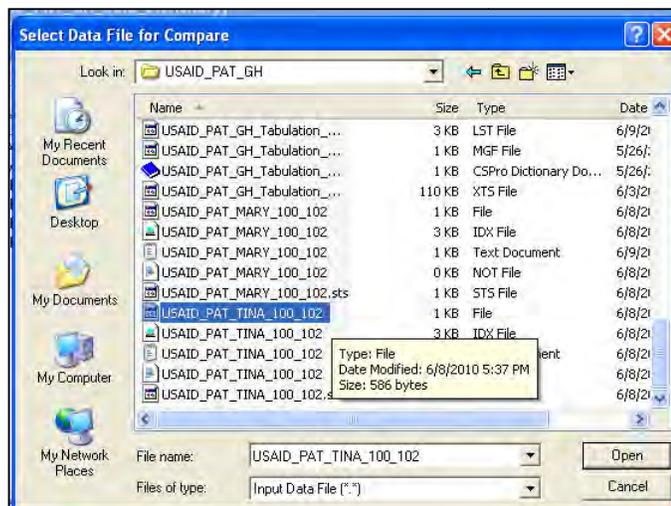


7. In the file menu, select the first of the two input data files you want to compare (“USAID_PAT_GH_Mary_100_102”) and click **Open**.



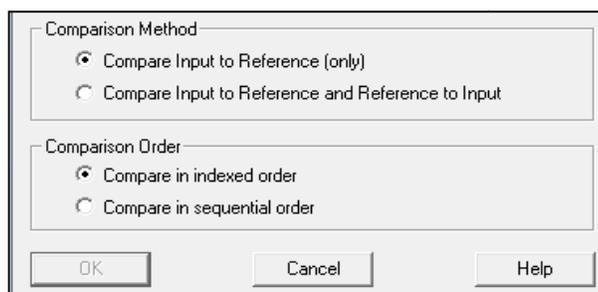
8. Click on the grey box to the right of **Reference File:** and select the second file (“USAID_PAT_GH_Tina_100_102”).

You can ignore the box next to **Listing File.**

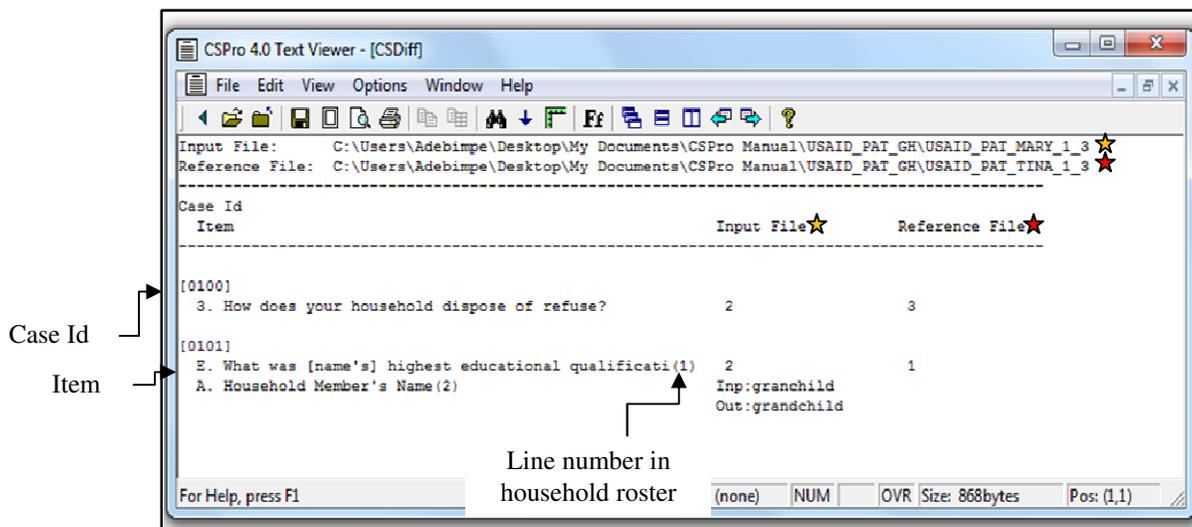


9. Under “Comparison Method”, click the button next to **Compare Input to Reference.** Under “Comparison Order” click the button next to **Compare in indexed order** and click **OK.**

IF there are differences between the two files, they will appear in the **Text Viewer** window.



The following image describes the contents of the **Text Viewer – [CSDiff]** window.



To save the output that shows the differences, select **File** at the top of the screen. In the **File** menu select **Save As**. Name and save the file in the folder with the Data Entry Template.

Follow the instructions for modifying the data on page 86 to correct any errors you find, referencing the Case Id (survey number) and Item (question number) you need to modify. If there is any doubt as to what a correct value should be, go back to the paper survey and verify what is written.

When you close the program and are asked if you want to save changes, click **Yes**.

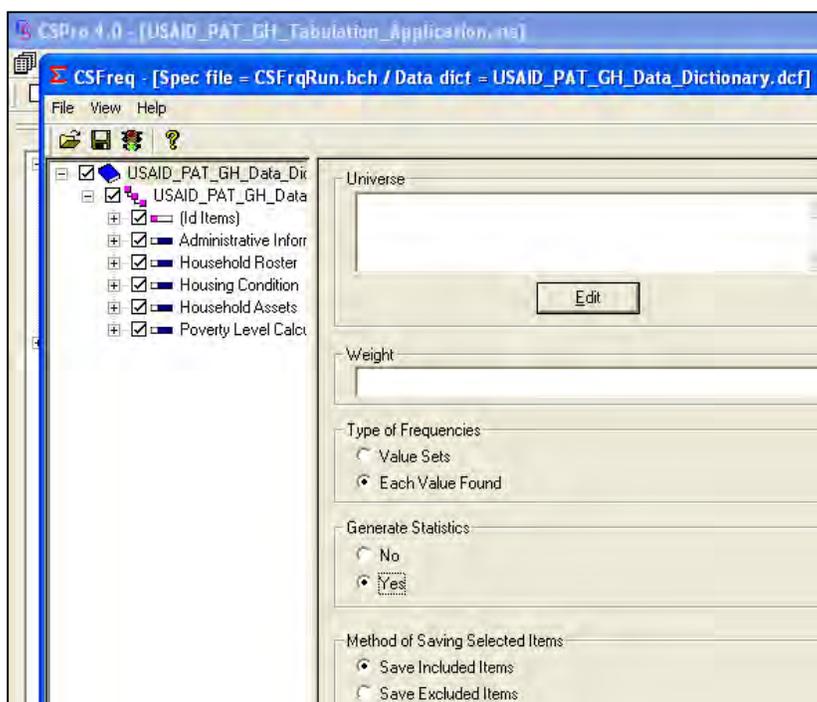
Once you have corrected the data, rename³⁷ one set of files to indicate it has gone through the **Compare Data** check (“USAID_PAT_GH_100_102c” – we use “c” to indicate the file has been *compared* with another). This file will be used to move forward with other quality control checks and eventually data analysis. The other file can be stored and left aside.

Tabulate Frequencies Tool (Looking For Odd Values)

Survey errors made by Interviewers (or bad handwriting!) can sometimes be found even after surveys have gone through a number of quality control checks. For this reason, we run “frequency” calculations on our databases to look for survey answers that just do not look right. An example would be a household that is said to own 80 of something when most others own closer to 8 of an item.

1. Open the **Data Entry Dictionary**. It is the file with the book icon next to it in your data folder.
2. Select the **Tools** menu near the top of the screen. Select “Tabulate Frequencies.”
3. Double click on the data dictionary file, “USAID_PAT_GH_Data_Dictionary.”
4. In the **CSFreq** window, select the variables for which you would like to generate frequency statistics. If you want to compare all variables (likely), click on the box next to the dictionary icon to select all.

³⁷ Re-name it in the folder in which it’s stored by highlighting the input data file and the four associated files. Right-clicking the input data file name and select “Rename.”



6. On the right side of the **CSFreq** window, select Type of Frequencies: “Each Value Found”, Generate Statistics: “Yes”, and Method of Saving Selected Items: “Save Included Items.”
7. Click on the stoplight icon in the toolbar. The **Select Data File(s) to Tabulate** window will appear on the screen.
8. Double click on the data file you wish to use (“USAID_PAT_GH_100_102c”).
9. The frequencies are presented in the **CSFrqRun** window. Click on the name of an individual question to see the frequency of responses.

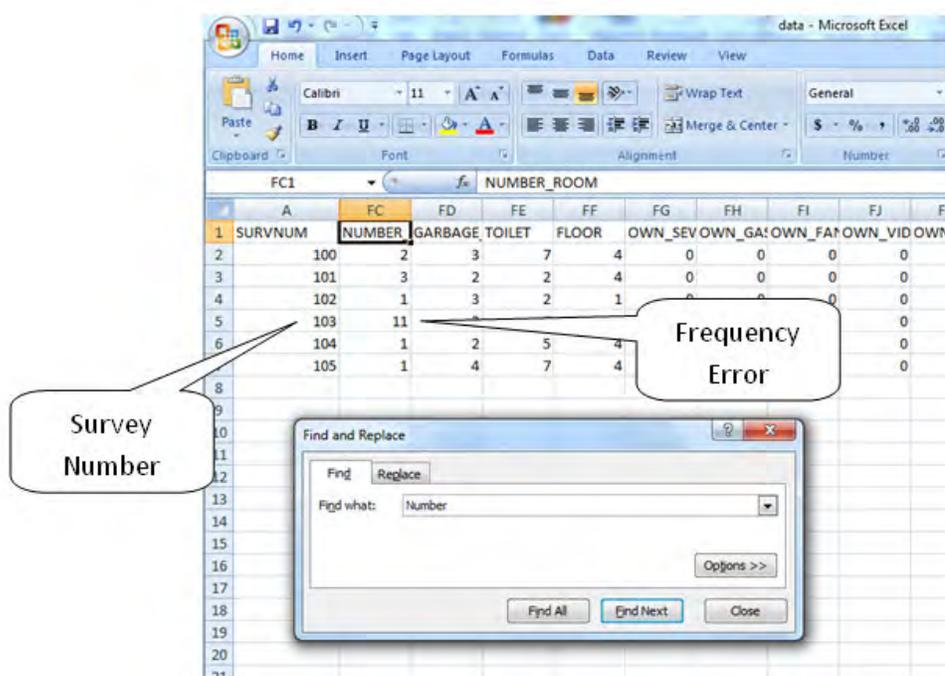
	Frequency		Cumulative	
	Total	Percent	Total	Percent
Total	3	100.0	-	-
1	2	66.7	2	66.7
11	1	33.3	3	100.0

Statistics : 2 numeric categories, Min 1 , Max 11 , Mean: 4.33 , Std.Dev: 5.7735 , Variance: 33.3333 , Mode: 1.0 , Median: 0.0000

10. In the example shown, it seems suspicious that one household has eleven rooms when all of the other households live in one-room homes. We need to identify the survey (case) in which this value occurs, verify on the hard copy whether or not a data entry or other error has occurred that can be resolved, and then make the correction.
11. To find the survey in which an outlying value occurs you will need to export the data set and search for the suspect value as follows.

Correcting Surveys with Incorrect Frequencies

1. Follow the directions in the “Export Data Tool” section of this chapter. Export to CSV.
2. Open the CSV document in Excel.
3. Find the column for the variable containing the suspect value either manually or by using the “Find” feature.
4. Identify the survey number of the case containing the error.



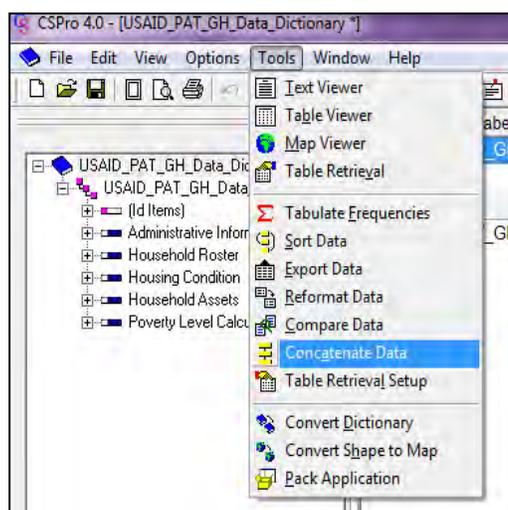
5. Identify the correct value.
6. If necessary, correct the error manually following the steps in the “Modifying Data and Adding Cases” section.

7. If the correct value is unclear, make note of the problem, notify a supervisor, and find the relevant case using the “Data Entry Application.” Delete the suspect value and save the case as a “Partial Save” so that it is not included in the poverty calculation until a correct value has been entered.

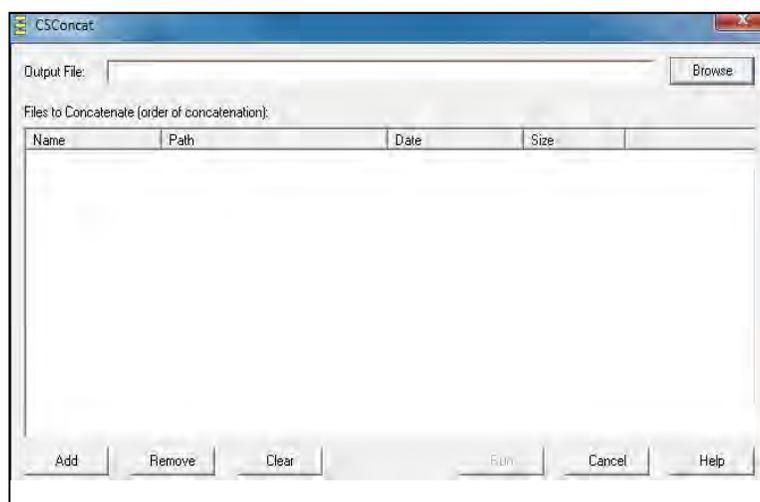
Merging Data (Concatenate Data Tool)

If different surveys are entered into two different data files, they will need to be merged. The **Concatenate Data** tool is used to perform this function. In this example we will be merging “USAID_PAT_GH_100_102cf” and “USAID_PAT_GH_103_105cf” into one new file which will contain survey numbers 100-105.

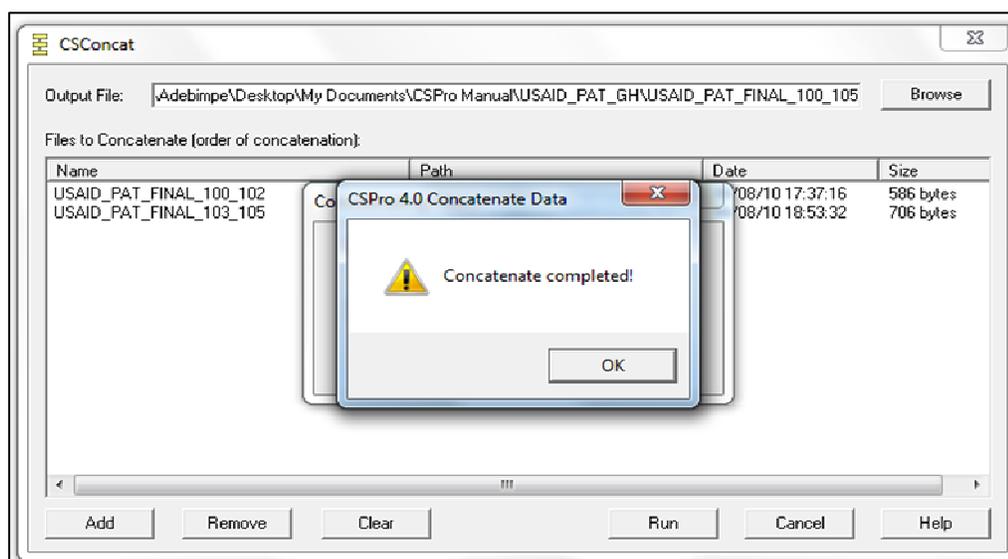
1. Open the **Data Entry Dictionary**. It is the file with the book icon next to it.
2. Select the **Tools** menu near the top of the screen. Select “Concatenate Data.”



3. The **CSConcat** window opens.



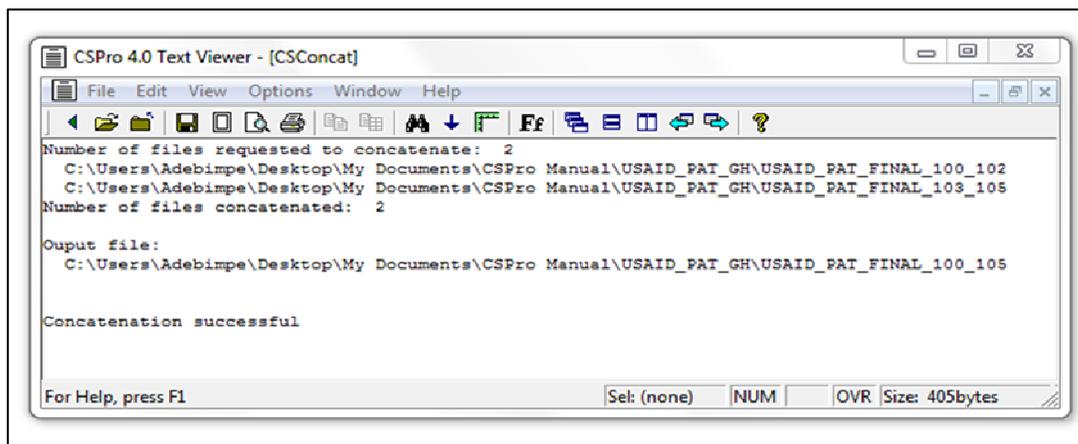
4. The output file is the file that receives the results of the concatenation (the file that contains the merged data). Press the “Browse” button and name an output file into which separate databases will be merged³⁸. We will name this output file “USAID_PAT_GH_Final_100_105” because we are merging “USAID_PAT_GH_100_102cf and “USAID_PAT_GH_103_105cf.” In the “Output File” box enter “USAID_PAT_GH_Final_100_105.”
5. Click “Add.” The **Select Files to Concat** window opens. Double click on “USAID_PAT_GH_100_102cf.”
6. Click “Add” again. Double click on “USAID_PAT_GH_103-105cf.” The order of the list of files to concatenate is the order of concatenation. To change the order of concatenation you need to rearrange the file names in the list. You can rearrange the names by drag and drop or sorting.
7. When you see the files you wish to merge listed, click “Run.”³⁹ When the merging is complete, a box will say “Concatenate completed!”



8. Then, the CSPPro Text Viewer will open a window that contains a summary of the merging process. CSPPro will give you a message indicating whether the merging process was successful or unsuccessful.

³⁸ It will look as if you should be searching for a file but instead, you should type in a new file name. Merging into an existing file will over-write its contents.

³⁹ It is possible to merge more than two files together into one document. Continue using the Add button until you have added all the files you want to merge into one list.



Tabulating Calculation of the Poverty Level

The poverty calculation program is run as data is entered into the Data Entry Application. The results of the poverty calculation program are summarized using the Tabulation Application. To summarize the results of the poverty calculation program, complete the following steps.

1. Open the Tabulation Application by double clicking on the CSPro **Tabulation Application** file. It is the file with red box next to it (see the figure below).



2. Click on the stoplight icon in the toolbar. The **Define Tab File Associations** window will appear on the screen.
3. Click on the grey box to the right of "<Input Data>." Double click on the final data file "USAID_PAT_GH_Final_100_105." Then click "OK."

Table 1. Poverty Status by Poverty Line

For reference purposes (using the same survey data from 2005 used to construct the PAT): In Ghana, 11% of households were living below the \$0.75 a day poverty line, 17% of households were living below the \$1.00 a day poverty line, 24% of households were living below the \$1.25 a day poverty line, 44% of households were living below the \$2.00 a day poverty line, and 55% of households were living below the \$2.50 a day poverty line.

Note: Poverty lines at 2005 PPP.

	Total
Poverty Status for \$0.75/day/PPP	
% Households above line	100.0
% Households below line	-
% Total	100.0
Households above line	6
Households below line	-
Total	6
Poverty Status for \$1.00/day/PPP	
% Households above line	100.0
% Households below line	-
% Total	100.0
Households above line	6
Households below line	-

For Help, press F1

4. The results show the percentage of clients living above and below a certain poverty line. The result for Microenterprise Results Reporting (MRR) is highlighted in yellow. In this example, 0% of households live below the \$0.75 a day poverty line; 0% of households live below the \$1.00 a day poverty line; 0% of households live below the \$1.25 a day poverty line; 16.7% of households live below the \$2.00 a day poverty line; and 33.3% of households live below the \$2.50 a day poverty line. It should be noted that the model is calibrated to both the \$1.25 and \$2.50 lines⁴⁰, but not the other lines. The results at the other lines are likely less accurate and should be used for indicative purposes. The table also shows, for reference purposes, the percentage of households living below different PPP poverty lines for the survey data used to construct the PAT.

⁴⁰ For PATs that use the median and national poverty lines, the model is calibrated to those.

Table 1. Poverty Status by Poverty Line

For reference purposes (using the same survey data from 2005 used to construct the PAT): In Ghana, 11% of households were living below the \$0.75 a day poverty line, 17% of households were living below the \$1.00 a day poverty line, 24% of households were living below the \$1.25 a day poverty line, 44% of households were living below the \$2.00 a day poverty line, and 55% of households were living below the \$2.50 a day poverty line.

Note: Poverty lines at 2005 PPP.	Total
Poverty Status for \$1.25/day/PPP	
% Households above line	100.0
% Households below line (MRR)	-
% Total	100.0
Households above line	6
Households below line	-
Total	6
Poverty Status for \$2.00/day/PPP	
% Households above line	83.3
% Households below line	16.7

5. Scroll down to see the results for more poverty lines. The results to be used for MRR reporting will be highlighted.

Modifying the Tabulation Calculator

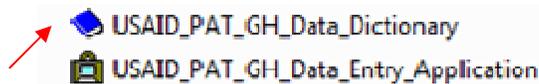
It is possible to alter the tabulator tool to view alternative tabulations using different variables. Please visit the CSPro User's Guide Version 4.0, available at <http://www.census.gov/ipc/www/cspro/doc.html>, for more details. The section covering the Tabulation Application begins on page 163.

Export Data Tool

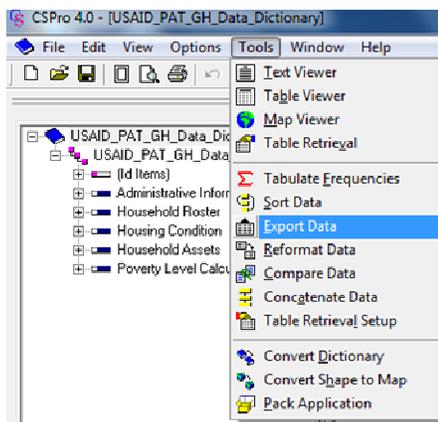
The "Export Data" Tool is a tool that you may find useful in analyzing your survey data. The tool allows you to export data from a CSPro input data file to tab- or comma-delimited text files that can be imported into spreadsheets or databases. For easier import into spreadsheets and databases, household and individual level data should be exported separately.

To export household level data complete the following steps.

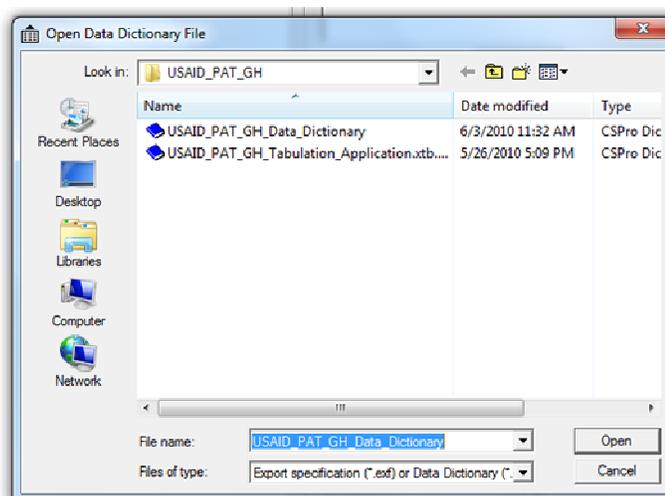
1. Open the Data Dictionary by double-clicking on the CSPro Data Entry Dictionary. It is the file with the book icon next to it in your data folder.



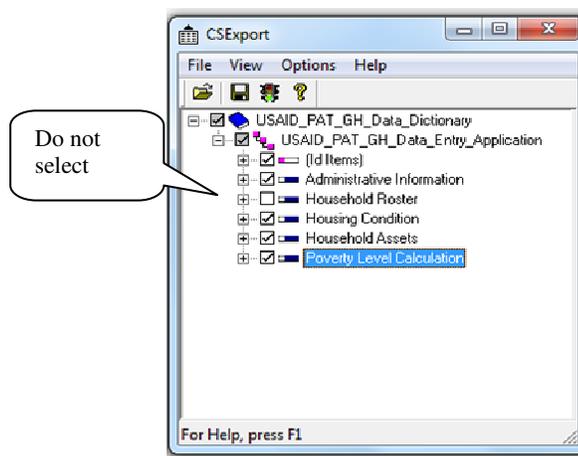
2. Select the “Tools” menu near the top of the screen, then “Export Data.”



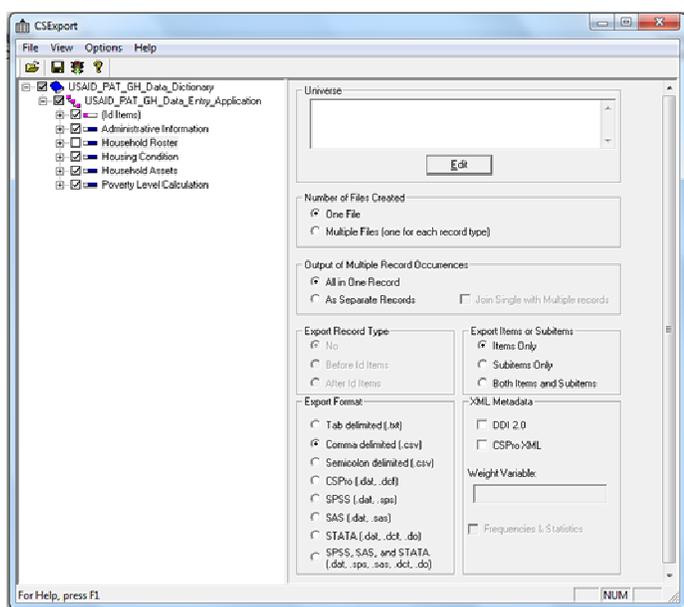
3. In the file menu, select the dictionary file and click “Open.”



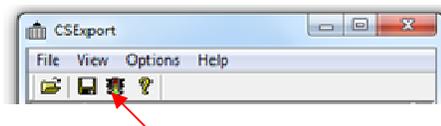
4. In the “CSExport” window, click every box in the data dictionary tree. Note that including all the variables under “Household Roster” will export each individual family member’s name. Use the + sign next to the check box to expand the list of individual variables and select only those you wish to include.



5. In the “CSExport” window, click “One File”, “All in One Record”, and “Comma delimited (.csv)” (or another format as you like).



6. Click on the stoplight icon in the toolbar. The “Select Data File(s) to Export” window will appear on the screen.



7. In the file menu, select the data file you want to export and click “Open.”

8. In the “Specify Name of Exported File” window, enter the file name and click “Save.”

9. The “Text Viewer” window will open, and you can close the window.

10. Note: To open the file in Excel, first open the Excel program, select Open from the main menu, and open the saved file. If you do not see it listed, change “Files of Type:” to “All Files.”

Reporting to USAID

If USAID requires your organization to report the percentage of your institution’s clients who are “very poor,” then the final database used to do the poverty calculation should be sent to the PAT Help Desk along with your results and a copy of the survey used in the field, upon completion. When sending the database, make sure to use WinZip or another file-compression software package to compress the database file prior to sending it to the Help Desk as an attachment. Subsequently, you will include the results from your PAT implementation with your regular online MRR submission. If you are required to measure poverty and report to USAID, and are having trouble with the data processing or any other aspect of the implementation process, please contact the PAT Help Desk at pathhelp@iris.umd.edu.

CHAPTER SEVEN – ADDITIONAL RESOURCES

PAT Website: www.povertytools.org

The www.povertytools.org website provides links for downloading all of the USAID-certified tools, corresponding Data Entry Templates, country-specific User Guides, and the latest version of this PAT Implementation Manual, as well as a list of frequently asked questions (FAQs) to assist you in learning the procedures for tool implementation and meeting USAID reporting requirements.

Online Training

A full course in PAT implementation is available online at www.povertytools.org/PATtraining. It is a self-paced course, moderated by PAT Trainers. It discusses all aspects of understanding and using a PAT and includes activities, printable resources, self-assessments and graded activities. Users are also led through the process of developing a PAT implementation plan throughout the course and will have a useable, complete plan at the end of the course if all instructions are followed. PAT trainers provide feedback and support throughout the course. Certificates are awarded to participants who complete at least 85% of the course materials successfully.

PAT Help Desk: pathelp@iris.umd.edu

The PAT Help Desk is available **only** for those organizations that are required to implement a PAT. If you are required to implement a PAT as part of USAID reporting requirements, and have any questions about the tools and implementation process, send a message to the address above with:

1. Your name
2. Name of your organization
3. Country where you operate
4. Name of USAID-funded project with microenterprise activities
5. A detailed description of the issue and the pertinent questions you have.

Please make sure that the response from the PAT Help Desk does not end up in your spam or junk email folder.

Help Desk support can also be provided via phone or Skype after being arranged by email.

Regional Trainings

Currently, four-day PAT trainings are held each spring in Washington, DC. Registration preference is given to required implementers, though there are typically spots open for others interested in learning about the PAT. These trainings are free, though participants must pay for any travel, lodging or meal expenses they incur (with the exception of daily lunch).

With sufficient demand, IRIS will also host regional trainings outside the US and has previously conducted them in Peru, Cambodia, Uganda, and Ethiopia. Contact pathelp@iris.umd.edu to request a training in your region. Note that regional trainings will be offered ONLY if there is sufficient demand from USAID Missions and implementing partners.

APPENDIX A – IMPLEMENTATION TEAM JOB DESCRIPTIONS

Project Manager

The Project Manager will take overall responsibility for planning and implementing the field survey. The manager will oversee and participate in the training of staff, supervise the translation of the survey, and verify that the sampling is being done correctly and that only sampled clients are being interviewed. They will monitor progress towards completing the survey and will do “spot checks” to verify that Interviewers are following the questionnaires consistently during interviews and filling in the forms correctly and completely, and that the Field Supervisors are carrying out their quality control function. The Project Manager will monitor the team’s progress in staying on schedule and within budget as the field work progresses. Ideally, the Project Manager will have previous assessment or monitoring and evaluation experience, as well as an excellent track record for successfully managing resources and personnel.

The Project Manager should complete the following tasks:

<p><u>During Planning/Training</u></p> <ul style="list-style-type: none"> - Hire/assign team members - Plan, implement, and oversee training with Field Supervisor - Develop implementation plan (schedule, logistics, etc.) - Qualitative review of training materials, sample, questionnaire, and internal control system - Questionnaire translation and editing 	<p><u>During Interviewing</u></p> <ul style="list-style-type: none"> - Communicate with Field Supervisor on a regular basis - Inquire on a regular basis about results vs. targets - Conduct spot checks
<p><u>During Data Processing and Analysis</u></p> <ul style="list-style-type: none"> - Inquire about data entry issues and results vs. targets - Conduct spot checks 	<p><u>Throughout</u></p> <ul style="list-style-type: none"> - Budget - Communicate with Help Desk, as needed - Submit deliverables

Required Qualifications/Competencies

- Project management experience
- Staff management experience
- Monitoring & evaluation experience
- Budget management experience
- Good communications skills
- Team spirit/motivates staff
- Capable of delegating responsibilities
- Ability to translate final objective to field staff
- Knowledge of local context
- Good report writer

Field Supervisors

It is recommended that Field Supervisors manage between three and four Interviewers each. Only one Field Supervisor may be used if the travel logistics permit the use of one interview team for the field work. They should help with the interviewer training and translation of the survey. Field Supervisors are responsible for coordinating the daily activities of the Interviewers, including arranging movement to and from interviews and transport from one survey site to the next. They also take responsibility for strict quality control, including ensuring that the questionnaires are filled out correctly and completely and that the information contained in them is accurate before leaving each survey area. Field Supervisors check the work of each interviewer on a daily basis to minimize the number of errors and missing values. During the first few days of field work, a Field Supervisor may want to observe several interviews being carried out by those Interviewers deemed to be the least experienced or capable. Field Supervisors should also conduct occasional random spot checks to verify the accuracy of data by partially repeating a client or beneficiary interview without the interviewer being present.

Field Supervisors report regularly to the Project Manager and Sampling and Survey Tracking Coordinator on progress, costs incurred, and any irregularities in the field, especially those that caused them to deviate from the sampling plan. They should have prior experience in conducting quantitative surveys, strong leadership skills, excellent attention to details and time management, and be assertive in supervising Interviewers to ensure that high-quality data is collected.

Field Supervisors should complete the following tasks:

<p><u>During Planning/Training</u></p> <ul style="list-style-type: none"> - Assist with translation of survey - Plan fieldwork logistics (transport, schedules, distribution of Interviewers) - Assist with and lead sessions during interviewer training 	<p><u>During Interviewing</u></p> <ul style="list-style-type: none"> - Coordinate daily activities of the Interviewers - Quality control (questionnaires filled correctly, completely, random spot checks) - Check/observe interview process - Report to Project Manager and SSTC (progress and irregularities)
<p><u>During Data Processing and Analysis</u></p> <ul style="list-style-type: none"> - Resolve problems brought up by Data Processors 	

Required Qualifications/Competencies

- | | |
|--|--|
| <ul style="list-style-type: none"> - Prior experience in conducting quantitative surveys (including as an interviewer) - Strong leadership skills (assertive in supervising Interviewers; listening skills) - Attention to detail | <ul style="list-style-type: none"> - Time management - Ability to receive criticism/to provide constructive criticism - Initiative and ability to make decisions and solve problems on the spot |
|--|--|

Interviewers

Interviewers with prior field and survey experience are desirable, but just as important are individuals with strong communication skills who can carry out interviews in a confident and relaxed manner while maintaining consistency. All Interviewers require thorough training that includes in-depth review of the questionnaire to understand its intent and repeated practice in posing the questions in the local language.

Do not send field staff to interview clients with whom they have had previous business or personal relationships. Institutions have the option to use administrative staff or to use field staff from other regions so that they do not interview anyone they know.

Interviewers should complete the following tasks:

<u>During Planning/Training</u>	<u>During Interviewing</u>
<ul style="list-style-type: none"> - Participate in all training activities - Understand the aim of the survey - Review and learn the questionnaire - Repeat practice in interview role playing, using the questionnaire - Help to field test the questionnaire and discuss how the process and translation might be improved - Be available throughout the training process 	<ul style="list-style-type: none"> - Respect the schedule and planning process - Respect clients and explain questionnaire according to their level of education and ability to understand - Follow instructions - Ask questions with precision - Record information accurately - Use probing questions when necessary - Maintain the same quality of work throughout the process - Resolve any logistical or other problems or report to the Field Supervisor as necessary - Take an active part in the de-briefing process - Be aware of process issues

Required Qualifications/Competencies

- Knowledge of local language
- Previous field and survey experience
- Observation skills
- Problem-solving skills
- Willingness to follow instructions and accept supervision
- Precision in asking and recording information
- Good disposition and ability to put the interviewee at ease
- Efficiency in getting the interview done while meeting all the cultural norms for politeness, thus leaving a good (or at least neutral) impression about the implementing organization in the mind of the interviewee

- Curiosity about the research topic that will lead to better probing, ensuring that they are recording what the interviewee said and not what they expected that the person “should say”
- Culturally acceptable appearance
- Willingness to work long hours while maintaining the same quality in the last interview as in the first interview
- Ability to observe body language, the quality of house construction, the setting in the house, the tools and stock in the business, etc., and to use probing questions when the answers seem to contradict what is being seen

Sampling and Survey Tracking Coordinator

The Sampling and Survey Tracking Coordinator plays a key role in tracking the progress of survey implementation and maintaining the sampling plan. Experience in sample design is critical and might require working with a consultant or other technical expert to help fill this role.

The Sampling and Survey Tracking Coordinator should complete the following tasks:

<p><u>During Planning/Training</u></p> <ul style="list-style-type: none"> - Work with Project Manager and Data Processing Coordinator to set up numbering system for all surveys: test surveys and actual surveys - Participate in sampling selection and oversee the assignment of survey numbers. - Set up a system to monitor location and status of each survey, including the assignment of surveys from the replacement list - Assist with training, as needed 	<p><u>During Interviewing</u></p> <ul style="list-style-type: none"> - Monitor distribution and collection of surveys from each sampling group, including constant communication with all Team Leaders and Project Manager (or Coordinator) to ensure that all production goals are being met - Assist with quality control of surveys and returning of surveys to the field, if necessary - Provide daily feedback to Field Supervisors and Project Manager so that problems can be corrected quickly - Assist with supervision of data processing and other duties, as needed
<p><u>During Data Processing and Analysis</u></p> <ul style="list-style-type: none"> - Confirm that all surveys have been handed in and are completely filled out. - Compile statistics totaling number of surveys completed for each sampling group and supervise a physical counting of documents to confirm the totals. - Assist with data cleaning and analysis, and with all tasks involved in generating statistics, graphs and tables for the final report. - Assist in final organization of computer files and filled-in questionnaires. 	

Required Qualifications/competencies

- Experience in quantitative sampling techniques
- Field research experience
- Attention to detail
- Team management experience
- Ability to make quick decisions

Data Processing Coordinator

The Data Processing Coordinator oversees the data processing system. They train the Data Processors and perform spot checks to ensure data is accurately entered into the system.

The Data Processing Coordinator should complete the following tasks:

<p><u>During Planning/Training</u></p> <ul style="list-style-type: none"> - Lead training on Epi Info or CSPro, survey coding and data entry techniques - Participate in survey design and how that affects the data input and forms - Install Epi Info or CSPro software and copy all pertinent files containing forms and procedures to hard drive of all applicable computers - After supervising translation of Data Entry Template into local language (if needed), personally test data entry - Directly oversee data entry of first few surveys to ensure everything is being done correctly 	<p><u>During Interviewing and Data Processing</u></p> <ul style="list-style-type: none"> - Participate in and oversee quality control of survey forms, and make sure that incomplete forms are returned to appropriate supervisors for completion - Solve software problems - Supervise data input and cleaning - Oversee file management, file merging, and maintaining backup copies of databases - With other team leaders, set up a standard format for distributing updates and corrections to all members of evaluation process (such as a bulletin/newsletter)—and contribute information daily
<p><u>Finalizing Data Processing and Analysis</u></p> <ul style="list-style-type: none"> - Ensure that all surveys have been entered and re-entered, and that appropriate error-checking functions have been run. - Data cleaning to ensure all records are complete and properly entered. - Meet with Project Manager or other executive management to discuss specific goals for data analysis (if further analysis is carried out) - Analyze data, including demographical information - Assist all personnel working on final report in the retrieval of data, analysis, and creation of tables and graphs for presentation of data - Ensure survey forms are filed and kept in a safe and accessible place, and that final database is properly backed up. Both the survey forms and database need to be kept for a possible audit by USAID 	

Required Qualifications/Competencies

- | | |
|--|---|
| <ul style="list-style-type: none"> - Professional degree or extensive experience in data processing - Experience in data analysis - Ability to ensure confidentiality | <ul style="list-style-type: none"> - Leadership capabilities - Teamwork - Good communication - Decision-making initiative |
|--|---|

Data Processors

Those entering the survey data into the computer should have experience in both data entry and related quality control procedures. It would be preferable for them to have experience in the “cleaning” of statistical data and data entry. The ability to create graphs in Excel and also use PowerPoint would help in the preparation of any presentations to be given to the management, Board or USAID.

In many cases, personnel involved in field operations may be the same as those who later participate in the data analysis. Even those with no statistical training can help those who do have it in interpreting what the results mean for your particular organization.

Data Processors should complete the following tasks:

<p><u>During Planning/Training</u></p> <ul style="list-style-type: none"> - Receive specialized training - Help create logistical plan, including software and equipment needs - Learn to use Epi Info or CSPro and practice data entry - Test computers and software 	<p><u>During Interviewing</u></p> <ul style="list-style-type: none"> - Finalize all equipment and systems (1st day) - Daily data entry - Back up files on a daily basis - Report problems/obstacles to Data Processing Coordinator
<p><u>During Data Processing and Analysis</u></p> <ul style="list-style-type: none"> - Continue data entry - Quality control of your own work - Report problems to Data Processing Coordinator - Provide Data Processing Coordinator with requested information 	

Required Qualifications/Competencies

- Two-three years experience in data entry (statistics a plus)
- Knowledge/experience with MS Word, Excel, PowerPoint (depending on need)
- Detail oriented
- Availability throughout project period
- Ability to keep data confidential

APPENDIX B – GUIDE FOR INTERVIEWER DEBRIEFING DISCUSSIONS

Purpose of this discussion: This exercise will help the trainer ensure that each interviewer understands how to conduct the survey accurately and respond to unforeseen complications properly.

Participants: Interviewers after they have had experience in asking the PAT questions to the clients.

Personnel needed: A facilitator should ask the questions and an assistant should be available to write down all that is said relating to each topic.

When to be used: Discussion groups should be held during interviewer training, after successive rounds of pre-testing / practice interviews.

Instructions: The instructions to the facilitator are given in italics in the table below. Remember to encourage honest and open critique and comments in order to help the Interviewers correct problems that they have encountered. Please encourage suggestions about how the translation and process could be improved for your implementation and in the future.

Use of the following table: In the first column are the core questions, which are the main questions that the facilitator is asking. In the next column to the right are the probing questions to go along with the core question.

<u><i>Core Questions</i></u>	<u><i>Related Probing Questions</i></u>
<p><u>Warm-Up Questions Regarding the Process</u></p> <ol style="list-style-type: none"> How are the interviews going? How many interviews were you able to do today? How long did each interview take? 	<ol style="list-style-type: none"> What are the greatest difficulties you are having? Are you having difficulties finding the respondents? What is going well? What are the main factors affecting the time it takes you to complete an interview?
<p><u>Training Questions</u></p> <ol style="list-style-type: none"> Do you feel that you were given enough training to properly implement the tool? What were the most useful parts of the training? What were the least helpful parts of the training? How could it have been improved? 	<ol style="list-style-type: none"> Why? Why not? List specific parts and why they were useful. List specific parts and why they were not useful. How should it be done differently in the future?
<p><u>Translation of Questions</u></p> <p>Does it read well in your language?</p>	<p>How could this be improved? How do you make sure you are really asking the same question the same</p>

<p>Please take out a survey form. We are going to look carefully at each of the questions and see where you may be having difficulties. We will start with the questions about the client and household. <i>The first time you hold this discussion, you should go question by question. The next times, you can go section by section.</i></p> <ol style="list-style-type: none"> 1. Question 1 2. Question 2 3. Question 3 Etc. 	<p>way to each respondent? <i>For each question:</i></p> <ul style="list-style-type: none"> • Did you have any problems interpreting to the clients the meaning of this question (or the questions in this section)? • Which questions did the clients have difficulty in understanding? Why? • <i>If it is the first or second day</i>, what are other ways to explain the question so it would be more easily understood? • Do you think that the clients are manipulating their answers? Why? • Is there a way that we can verify this information? Can we observe contradictions in the household or in the data that this client has provided to us on other forms? • Were there any questions that clients refused to answer? Why do you think this was so?
<p><u>Building Trust with Clients to Provide Correct Answers:</u></p> <ol style="list-style-type: none"> 1. What are the techniques you are using to build the trust of the clients so that they give you answers that are reasonably accurate? 2. How do you assure clients when they express concern about why you need this information? 3. Did you find that clients would intentionally misreport their answers? How was this being done? 4. What could be changed to improve their willingness to give us better answers? 	<ol style="list-style-type: none"> 1. What techniques should we teach in future trainings? 2. What would you recommend that we say to clients? 3. Why do you think the clients were doing this? 4. Would it be better if we interviewed clients in a different place? Would it be better by making sure that other household members were not present? What could be changed to improve their willingness to be frank?
<p><u>Amount of Time:</u></p> <ol style="list-style-type: none"> 1. Did the clients make any remarks about the amount of time it took to answer questions? 2. How do you think we could make the process more efficient in terms of time? 	<ol style="list-style-type: none"> 1. What remarks did they make? 2. What specific suggestions do you have?
<p><u>Final Questions</u></p> <ol style="list-style-type: none"> 1. Are there any changes that we should make in the process of interviewing? 	<ol style="list-style-type: none"> 1. What specific changes?