Chapter 2

Research Design and Field Work Methodology

W. Timothy Farrell and Frederick L. Bates

Introduction

Research designed to evaluate human intervention programs or to examine hypotheses concerning social change and development requires a carefully thought out and executed plan. This is especially true if hypotheses concerning the roles played by various causal variables are going to be tested (Rossi and Freeman 1982:62). The basic methodological problem confronting such research is to devise a method whereby the changes observed over time can be attributed to the human intervention program rather than to other causes, particularly those which produce "normal" change trends in the system being studied (Rossi and Freeman 1982:38).

In any society or community, whether impacted by a disaster or not, change is constantly underway. In developing societies in particular, modernization trends are taking place and the societies are moving in one direction or another with respect to development objectives. Furthermore, in a country like Guatemala, development programs may be in the process of execution when the country is struck by a disaster and then affected by disaster relief and reconstruction programs. If the effects of disaster related programs are to be assessed, it is necessary to employ a research design which can separate the trends produced by on-going

change processes and pre-disaster intervention programs from those peculiarly associated with the disaster and its associated post-disaster intervention programs. If this is not done, then it will be impossible to tell which of the changes observed in the post-disaster period are truly disaster related and which are merely continuations of on-going processes (Campbell and Stanley 1966:13).

Because of these problems, research on social change and development associated with disasters which hopes to evaluate the effects of intervention strategies calls for an experimental design. (For description see Campbell and Stanley 1966:13, and Weiss 1972:60.) Such a design employs a control and experimental group along with before and after measures on relevant variables and characteristics. It furthermore assigns subjects (individuals or groups) to the experimental or control group randomly so that they represent unbiased samples of the same population. The experimental treatment or intervention is then introduced (in this case the earthquake and the disaster mitigation inputs) only into the experimental group, maintaining isolation of the control group from these interventions (Campbell and Stanley 1966, Weiss 1972).

If this design is adhered to, then the investigator can reasonably attribute changes in the experimental group beyond those observed in the control group to the intervention. If, however, any deviation occurs from this design, problems arise in interpreting results since there are a number of possible competing explanations for what is observed.

In the study of post-disaster reconstruction many of the conditions listed above are impossible to attain and others can only be approximated.

As a consequence, the best design which can be used is one which is only quasi-experimental since it only approximates the ideal experiment (Campbell and Stanley 1966:34, Weiss 1972:67, Rossi and Freeman 1982:217). In particular, it is impossible to achieve randomization in the assignment of units to the control and experimental group. Potential membership in these groups by individuals, households, or communities is determined by the disaster event and where it strikes. Those people and those households and communities stricken by the disaster become a pool from which an experimental group may be chosen. They are potential experimental group subjects because they experience the "experimental treatment," in this case the earthquake, and post-disaster relief and reconstruction inputs. Those not directly affected by either the earthquake or relief and reconstruction programs become a potential control or comparison group (Weiss 1972:69, Rossi and Freeman 1982:219).

For two particular reasons, however, this potential comparison group can only act as a "weak control" in experiment terms. First, potential membership is determined by the non-random effect of the earthquake and disaster mitigation programs. This means that the two sectors of the same society from which the control and experimental group are drawn may be quite different from each other to begin with. One might be changing at a faster rate than the other, or one may start out measuring higher or lower on some critical attribute, for example economic resources, than the other. As a consequence of such inequalities it will be difficult to separate differences in changes produced by the disaster and recovery process from those produced by inequalities between the two groups. The

experimental group (disaster victims) may change faster than the control group (non-victims) because they were more receptive to change to begin with and were already changing at a more rapid rate and not because of disaster related effects.

There is a second reason that "control groups" in disaster research designs executed in small developing countries like Guatemala are at best "loose controls." The affected area from which the experimental group is selected, and the unaffected area from which the control group is chosen are close to each other and effects of the earthquake may "spill over" and affect the control group. Thus the experimental "treatment" is not kept exclusively in the experimental group and because of this, part of the change in the control group must be attributed to the disaster. This means that it will take a larger change in the experimental group produced by the earthquake and reconstruction inputs to register as significant in statistical terms. As a consequence, disaster related change may actually occur and appear to be attributable to non-disaster change processes.

Because of these two difficulties, the best that can be achieved in most disaster research situations is a quasi-experimental design which uses a "weak control group" for purposes of comparison with an experimental group. Such a design was chosen for this study, with full knowledge of its limitations, since such a design is still superior to one which neglects any comparison with groups outside the immediate disaster area (Campbell and Stanley 1966:47).

There is still another reason that an ideal experimental design can

not usually be achieved in a disaster study. Such a design requires longitudinal data which measures key variables in both the experimental and control groups both before and after the experimental treatment, in this case the disaster and disaster related mitigation programs. Since disasters are seldom predicted in advance and since research funds are almost never available to gather pre-disaster data on communities or societies likely to be struck, pre-disaster measures of key variables with respect to the exact units which are later studied are not available. Studies begun after impact must therefore depend upon public data sources which never quite fit the needs of the researcher, or upon retrospective data compiled from the memories of victims and public officials. Such data introduce a source of error into the research process which is of unknown proportions and is difficult to overcome (Bates et al 1963:174-177).

Since data on the pre-disaster situation of both the control and experimental groups are collected using this method however, it is unlikely to produce different results with the two comparison groups and differences observed over time between them can safely rule out this factor as a source. If retrospective data introduces systematic errors it should have the effect of exaggerating or minimizing change in both groups rather than differentiating between them.

Some of the data obtained from memories of respondents can be checked against public records and published statistics, most of the time at the aggregate level, and judgements can be made as to whether they exaggerate or underestimate the true pre-disaster situation. Nevertheless, such

data introduce a potential source of error into the research process.

In this study such data were obtained through interviews with household heads and community leaders to establish the physical and economic resources of households and communities immediately before the disaster. These data pertain to such subjects as household composition, characteristics of the house itself and of the physical facilities it offered, the occupations and incomes of family members, their land ownership or land tenure situation, the production of agricultural products in the year preceding the earthquake, and so forth. These retrospective measures represent benchmarks against which change is measured subsequent to the earthquake. Also subject to retrospective methods were data collected on disaster relief and reconstruction inputs for the first one and a half to two years following the disaster. Beyond this time, data were collected contemporaneously on three time periods. It is these contemporaneous data that are compared to retrospective data to make change measures in this study.

The research design therefore can be termed "quasi-experimental" and "longitudinal" in that it employs an experimental and "weak control" group upon which measures are taken longitudinally, beginning with retrospective data and proceeding through three waves of data collection on current or contemporaneous situations. The broad outlines are given in Table 2-1.

Because this research is focused on the effects of disasters on social change and development and at the same time on how the characteristics of the sociocultural system affected by disaster respond to

Table 2-1

Characteristics of the Research Design

	Type of Data and Time Period				
	Retrospective Data		Contemporaneous Data		
Study Group	Pre-earthquake Time 1	Disaster Impact Time 2	2 yrs.After Impact Time 3	3 yrs.After Impact Time 4	4 yrs.After Impact Time 5
Experimental Group (Households in communities heavily damaged by the earthquake)	e^{X_1}*	e^{X_2}	e^{X_3}	e^{X_4}	e^{X_5}
Control Group (Households in communities lightly or unaffected by the earthquake)	c^{X_1}	c^{X_2}	c^{X_3}	c^{X_4}	c^{X_5}

* e^{X_1} - measure on a variable such as the value of the house occupied by a victim in the experimental group at Time 1, just before the earthquake. Numerical subscripts indicate same measure at succeeding time periods.

various forms of intervention, it was necessary to introduce additional sampling criteria to those implied by the selection of a control and experimental group. In particular, there were three dimensions of stratification introduced into the sample design. First, because of interest in the effects of cultural differences on disaster response, and because of interest in the cultural appropriateness issues and issues related

to the equity of distribution of aid, the sample was divided into an Indian and a Ladino sample. Guatemala is about evenly divided into these two ethnic groups with the Ladino group being in a dominant political and economic position. Both the experimental and control groups were therefore sub-divided into communities which were predominantly Indian and those which were predominantly Ladino.

Since these two populations are unevenly distributed geographically, with most predominantly Indian communities being primarily in the highlands, and most predominantly Ladino communities being concentrated in the East, an East-Highlands stratification was introduced into the sample along with the ethnic differences.

Finally, there was interest in looking at how community size, complexity and isolation affected the reconstruction development process since social organizational factors vary along these lines, and program design and delivery problems also are affected by them. It was decided therefore to stratify the sample according to the political status of the community in the Guatemalan governmental administrative system.

Guatemala is divided into departments, each of which has a departmental capital. These are next divided into municipios which are further subdivided into smaller places called aldeas. There is an even smaller unit called a casaria. Departments are like states in the United States, and municipios are like counties. Each has a central administrative center called a cabacera. It was this central unit which was selected for study. The control and experimental groups were then divided into department capitals, municipios and aldeas. Particular

units for inclusion in the sample were selected on a basis to be discussed later under the execution of a sampling plan.

The design, as discussed so far, excludes Guatemala City. This very large urban center, which serves as the capital of the country, was also struck by the earthquake and therefore fell naturally within the experimental group. However, there is no other city in the country which can be compared to it. It was therefore impossible to select a control group for comparison. Furthermore, the city had close to a million residents at the time of the earthquake and funds were not available to draw a truly representative random sample of the entire city, and at the same time collect data on towns and villages in the countryside. Since communities outside the city represent a variety of sociocultural organizational patterns, and since reconstruction programs of particular types were associated with particular communities, thus offering an opportunity for many cross-community, cross-program comparisons, it was decided to put most of the project's resources into data collection outside the city. There was the additional fact that development programs are concentrated there and the chance of observing the impact of disaster on development would be maximized by this procedure, given research funding limitations.

There was, however, a need to monitor a number of things going on in the city. In particular, reconstruction programs had been undertaken there to house disaster victims in newly formed neighborhoods. There were four types of situations known to exist. First, squatters settlements had grown up in various parts of the city and it was believed

these settlements were comprised mainly of disaster victims, many of whom had migrated to the city following the earthquake. Information was desired on the origin of these settlements and on their evolution as "communities" following the earthquake.

Secondly, the Guatemalan government had built refugee style housing settlements to take care of the large number of people who had moved into the streets and parks of the city right after the earthquake. Again, information was desired on the origin of these people, and on their eventual fate as the reconstruction process progressed.

A third type of urban housing area which grew up after the earthquake consisted of permanent houses built by means of agency programs to house disaster victims. This sort of housing development was believed to represent the final stage in the resettlement of squatters and victims who were housed in government disaster refugee centers. Such areas usually consisted of several hundred newly built detached houses and newly constructed community facilities and services and presented an opportunity to study the formation of a new urban neighborhood-community stemming from the disaster.

A final stype of unit was like the one just discussed, but was built to house people who were being resettled from a rural community which had been so badly damaged that it could not be fully rebuilt. This community represented one of the rare cases in which Indians were being resettled from rural areas into the city and presented an opportunity to observe the change processes associated with such a movement.

The city sample for this study therefore contains four urban neighborhoods of the types described above: (1) a squatters settlement,

(2) a refugee temporary housing project, (3) a newly built agency housing development for urban disaster victims, and (4) a newly built agency housing development for the resettlement of Indians from a heavily damaged rural community. The plan was to use these units as comparison groups for each other since no effective control group could be found for any of them.

The final sample design for this research at the level of the community is given in Table 2-2.

Table 2-2

Community Sample Design

Type of Community**	Experimental		Control	
	Indian	Ladino	Indian	Ladino
City	1*	3	0	0
Department Capital	1	1	1	1
Municipio	3	2	1	1
Aldea	4	4	2	1
TOTAL	9	10	4	3

* The community from which the Indians in the city came was also included in the sample. It was a municipio on the outskirts of the city and is included as an Indian municipio in this table.

**The East-Highlands division of the sample consists of Experimental, 6 East, 9 Highland; Control, 3 East, 4 Highland.

As can be seen from this table, the control and experimental groups are not balanced in the number of communities selected for study. It was decided because of the weak nature of the control to select only enough communities to provide a control for each of the classes of units in the experimental group so that comparisons could be made between department capitals, or municipios and aldeas, in both the Indian and Ladino categories when necessary. This permitted a larger sampling of the disaster area than would have been the case if a balanced sample of each had been used. Excluding the city and the one rural municipio associated with the resettlement of Indians, there are twice as many experimental group communities as control group ones. For a listing of the exact communities used in the sample and their classification according to sampling plan, see Table 2-3.

Sampling

Because the communities chosen for investigation were selected by a series of criteria other than strict probability sampling, e.g. design requirements, availability of pre-earthquake data, researchers' familiarity with the region, etc., it was of paramount importance that the sampling design used to select households for interview insure true representativeness insofar as possible.

Obviously one major problem to be overcome was the disparity in the size and kinds of units to be studied. How does one compare, for example, Chimaltenango, a large department capital, with Pacoc/San Marcos, a small divided aldea? Are the aldeas chosen truly representative

Table 2-3

Communities Selected for the Research

Experimentals	Loose Controls
Indian[1]	
Chimaltenango (Dept. Capital)	Sololá (Dept. Capital)
Patzún (Municipio)	San Lucas Tolimán (Municipio)
San Martin Jilotepeque (Municipio)	Cerro de Oro (Aldea)
Las Lomas (Aldea)	San Marcos La Laguna (Aldea)
San Marcos (de Puerto Rico) (Aldea)	
Pacoc (Aldea)	
Santa Maria Cauque (Aldea)	
Chinautla (Municipio)	
Ladino[1]	
El Progreso (Dept. Capital)	Cuilapa (Dept. Capital)
Sanarate (Municipio)	Barberena (Municipio)
Conacaste (Aldea)	El Junquillo (Aldea)
Santo Domingo Los Ocotes (Aldea)	
Espiritu Santo (Aldea)	
San Juan (Aldea)	
Zaragoza[2] (Municipio)	
Guatemala City	
Carolingia (Agency Housing Development)	
Roosevelt (Guatemalan Government Refugee Housing)	
4 de Febrero (Squatters Settlement)	
Nueva Chinautla (Agency Housing for Indians from Chinautla)	

of other aldeas supported by the same municipio? Does the weight given to the aldeas studied over or under-represent population under investigation? These and similar questions were critically asked throughout the process of elaborating the final sampling design. Ultimately it was decided that a modified multi-stage cluster design be used.

A cluster sampling design was chosen for use in this research and justified on the basis of several factors, not the least of which was cost. In the initial post-earthquake phases, not only was it impossible to "list" people, or families, but it was not even clear as to what was a house or a household. It was known that there were people "out-there," but there was no way of knowing how they were grouped and organized. In addition, individuals and families tended to be quite fluid during the early reconstruction period, living with friends and relatives, or alternatively accepting friends and relatives into their homes or temporary shelters. Thus, it was quite impossible to compile a family or household list or directory from which to draw a sample. In fact, that became a major section of the interview schedule itself.

A second factor in choosing a cluster design involved some well grounded assumptions about the areas and communities selected. Most of the communities were previously known to someone on the research team. Excluding Guatemala City, someone among the researchers had lived or worked in 18 of the 21 communities previously. Thus some assessment of the heterogeneity/homogeneity question could be made. As a consequence, two basic assumptions relative to the cluster design were formed.

1. Within smaller communities (rural) there is greater homogeneity than heterogeneity.

2. In larger communities,components (households) of a population are more or less systematically distributed (e.g. neighborhoods tend to share some common characteristics as manifested by the households that comprise them).

As a consequence, it was reasonable to assume that a few clusters chosen in small communities would be representative of the community as a whole. Similarly, it was assumed that given the method used to select clusters, representative data for large communities could be obtained. While it was recongized that cluster sampling may yield greater sampling errors than simple random samples of the same size (Blalock 1960:406). it was believed that the sample size and the longitudinal aspects of the research design would off-set this.

Mapping or "Listing"

A major concern in drawing the sample of households for this study was the fact that in the damaged towns, even the most current maps were rendered useless by the damage caused by the earthquake. Even in the control communities, the level of detail of the maps was inadequate for sampling purposes. As a consequence, all communities had to be re-mapped by the research team. In order to make maps sufficiently detailed for the purposes of this research it was necessary to visit every structure in all 26 sample units to verify if it was indeed a "house," and if it was in fact occupied. In addition, maps had to be highly accurate and clear so that the interviewers would be able to find the appropriate dwelling with a minimum of trouble; no mean task when confronted with labryinthine paths and house sites obscured from view by corn fields and coffee trees. Accurate maps were also essential since three waves

of interviews were planned with the same households. It was essential that interviewers be able to find the same house over and over again.

It must be noted that the mapping task was as complicated as it was crucial. Aside from normal mapping procedures, inquiries had to be made regarding whether individual structures were occupied. Do people live here? How many families use the kitchen? Is there another structure used by this family? The details on the map also had to be sufficient to permit interviewers to readily locate the structure and family. Even with the detailed attention paid to mapping, there were still problems in identification of the correct household and their dwellings when interviewing took place.

Sampling Procedures

Ultimately it was decided to aim for a ten percent sample of households in the communities selected. In some communities this would vary because of the small size of the village. A community, say, of 150 houses would only yield a sample of 15 and would be too small for any sort of within-community analysis. In the larger towns of Chimaltenango and Patzun a ten percent sample would be uneconomical and would have perhaps over-represented the households in these communities with respect to the total sample.

Using the maps, the communities were divided into sectors of approximately 20-25 dwellings each. This rule was overridden if the number was reasonably close and if there were some natural division such as a street, that made a more logical boundary. Next, the sectors were numbered continuously (throughout all communities) from 001 to 795

(the last sector numbered). Numbering was done in a serpentine fashion, criss-crossing each community with a "string" of continuous sectors. Thus, for purposes of sampling, a conceptually contiguous population was employed.

To determine the sampling interval for the selection of the sectors ten percent of the total number of houses was divided by five since it had been arbitrarily decided to randomly sample five houses in each sector. This number was then divided into the total number of houses, yielding the number of sectors to be chosen. For example, if there were 1300 houses in a community, a ten percent sample would be 130 houses. Since five houses for each sector would be sampled, it was thus necessary to draw 26 sectors (130 ÷ 5 = 26). Then, the number of sectors, in this case 26, was divided into the total number of houses, giving a sampling interval number. Thus, once the houses were grouped by sector, the interval would define the sector. Table 2-4 illustrates this procedure. Once the sector was defined, the houses in the sector were numbered 01 - N and five houses were selected from a random numbers table.

Table 2-4 summarizes the basic sampling system. However, the details - specifically the intervals used - sometimes had to be modified to suit local conditions. For example, in El Progreso it was found that the original interval calculated would probably not yield a 10 percent sample as required, especially if there were a large quantity of refusals, "not homes," and so on. Therefore the interval was reduced to 50.89. Also, as noted before, a decision was made to sample about 25 households, minimum, in the small communities. Therefore the sampling fraction in

Table 2-4

Example of Interval Sampling (Santa Maria Cauque)

Sampling Interval: 50.89

Sector (Cluster) No.	Number of Houses in Sector	Range	Selected Interval
00443	24	01 - 24	01.02
00444	21	25 - 45	No
00445	21	46 - 67	51.91
00446	21	68 - 87	No
00447	19	88 - 106	102.80
00448	22	107 - 128	No
00449	21	129 - 149	No
00450	21	150 - 170	153.69
00451	20	171 - 190	No
00452	24	191 - 214	204.58
00453	20	215 - 234	No
00454	20	235 - 254	No
00455	20	255 - 274	255.47

those communities is always larger than the predetermined ten percent. In addition, the communities of Pacoc, San Marcos Puerto Rico and Asentamiento Roosevelt were sampled somewhat differently. The interval system was the same (50.89), but because the houses in Roosevelt and San Marcos were "ordered," that is, formally arranged in neat rows, a simple system of sampling every sixth house was used. The interval system

indicated which block or row of house (i.e. cluster) would be sampled.

In Chinautla, the interval was reduced to about half, or 25.45. The rationale for this was the great number of empty houses due to migration and the high number of people employed daily in Guatemala City. Therefore, in order to guarantee an adequate sample, the lesser interval was chosen. As a consequence the original sampling fraction was 21 percent, or 72 houses. Even with this fraction, however only 45 interviews were ultimately obtained, yielding a final sampling fraction of 13 percent. Table 2-5 summarizes the sampling data from the first wave of interviews done in 1978.

As mentioned earlier in this section, both Patzun and Chimaltenango were problematic because of their size. Using the previously described system would have resulted in a sample of some 290 in Chimaltenango and 180 in Patzun, thus over-representing them for the purposes of this study. As a consequence, once these communities were mapped and sectors defined, about every third sector was eliminated, "reducing" the population (houses) by 24 percent in Chimaltenango and 32 percent in Patzun. The sampling system then proceeded as usual.

It will be noted that the city sampling unit of Asentamiento Roosevelt is seriously under-represented. This is due to a mapping error at the outset. The "Total Number of Houses" column represents the corrected total after the error was discovered,but the interviewing had been completed by then. Because of costs and time it was decided not to re-interview in that area.

Table 2-5

Sampling Data 1978 (EQ01)

Community	Total No. of Houses	Sampling Fraction Chosen	No.Houses Selected	No. Houses Interv.	% Int. Obtained	Sampling Fraction Obtained
Chimaltenango	2,022	.097	197	143	72.6	.071
Patzun	1,214	.099	120	107	89.2	.088
Zaragoza	871	.101	89	78	88.6	.089
San Martin Jil.	842	.095	80	66	82.5	.078
Las Lomas	65	.523	34	22	54.7	.338
San Marcos P.R.	88	.25	22	15	68.2	.170
Pacoc	48	.25	12	10	83.3	.208
Sta. Maria C.	294	.103	30	25	83.3	.085
El Progreso	967	.098	95	79	83.2	.082
Sanarate	1,278	.094	120	110	91.7	.086
Conacaste	198	.157	31	28	90.3	.141
San Juan	143	.189	27	23	85.2	.160
Sto. Domingo	203	.172	35	28	80.0	.138
Espiritu Santo	166	.169	28	25	89.3	.150
Cuilapa	877	.097	85	77	90.6	.087
Barberena	846	.095	80	50	62.5	.059
El Junquillo	131	.267	35	26	74.3	.198
Solola	1,061	.099	105	76	72.4	.071
San Lucas T.	738	.102	75	59	78.7	.079
San Marcos L.L.	171	.234	40	30	75.0	.175
Cerro de Oro	464	.097	45	31	68.9	.066
Carolingia	1,337	.09	120	101	84.2	.075
Roosevelt	1,870	.035	66	53	80.3	.028
4 Febrero	1,464	.099	145	117	80.7	.079
Chinautla	341	.211	72	45	62.5	.132
Nueva Chinautla	409	.159	65	49	75.4	.119
Average Sampling Fraction Obtained						.117

Health, Fertility and Nutrition Sub-Sample

A sub-sample of ten percent was selected upon which data related to health, fertility and nutrition were collected. This sample was obtained by drawing a random number from 1 to 10 to use as a starting point in each town. Then every tenth household was interviewed on these items.

Phase Two Sampling (EQ02)

In Phase II, the sampling universe was restricted to the damaged (experimental) communities only. It will be recalled that Phase II was designed to tap domains directly related to earthquake experience, thus the questions were irrelevant to members of the undamaged (control) communities.

Since a random sample as described above had already been drawn, and since households were being tracted over time, a simple convenience sample was drawn. A total of 256 households were interviewed. The communities were divided into "large" and "small" categories. In the "large" towns 32/33 households were selected and in the "small" communities the number varied from five to nine. Asentamiento Roosevelt and the Chinautlas were not sampled in Guatemala City since it was felt that 4 de Febrero was adequately representative of "squatter" settlements and Carolingia of "planned" settlements. Table 2-6 summarizes communities interviewed in Phase II.

Chimaltenango was not included in this wave of interviews. This was principally because of difficulty in obtaining interviews there due to a number of factors. Principally, many residents felt hostility towards agencies and outsiders in general because the town was used as a staging area and supply depot for many groups working in the area.

Table 2-6

Communities Interviewed in Phase II

Community	Number of Interviews
Sto. Domingo	8
Conacaste	8
Espiritu Santo	8
San Juan	8
Sanarate	33
Carolingia	32
4 de Febrero	33
San Martin Jil.	32
Las Lomas	9
Pacoc	6
Patzun	33
Sta. Maria Cauque	8
San Marcos P.R.	5
El Progreso	33

Residents felt that the aid was not being used to assist Chimaltenango which was severely damaged. In addition, because of its proximity to Guatemala City, numerous research groups, university students, and agency personnel, went to Chimaltenango to interview. By the time this study began over 50 waves of interviews had already taken place in this town and residents were hostile to interviewers. Because of these factors it was decided not to interview there in Phase II, recalling that it would be necessary to return for Phase III.

Criteria for inclusion in Phase II for individual households were:

1. Informant was a community leader either before or after the earthquake.

2. Informants were heterogeneous with respect to socioeconomic, ethnic and religious groups.

3. Informants were reasonably articulate (because of the nature of the Phase II schedules).

Thus informants (and alternates) were chosen before returning to the communities on the basis of the information obtained in the Phase I interviews.

Phase III (EQ03)

Phase III called for a 100 percent re-sampling of the Phase I population. No modifications were made in the sampling system. An attempt was made to revisit the communities in the same order as Phase I so that the time between interview waves was approximately the same for each community. Some minor changes were made because of weather problems (heavy rains make some towns nearly inaccessible at times), but generally the sequence was maintained.

Attrition was not as severe a problem as initially feared. Overall, the attrition rate from Phase I to Phase III was only 15 percent. Table 2-7 records the attrition rates for each community.

Summary

Briefly, then, the sampling system used was basically a cluster sample modified to meet local requirements. The Phase I sample was the basis for all subsequent samples and sub-samples. Convenience

Table 2-7

Attrition Rates Phase I - Phase III

Community	Phase I	Phase III	Attrition Rate (%)
Sto. Domingo	28	24	14.3
Conacaste	23	27	4.6
Espiritu Santo	25	22	12.0
El Junquillo	26	22	15.04
San Juan	23	21	8.7
Sanarate	110	92	16.04
Roosevelt	53	44	17.0
Carolingia	101	84	16.9
4 de Febrero	117	95	18.9
Nueva Chinautla	49	45	8.2
Chinautla	45	34	24.5
Chimaltenango	143	118	17.5
San Martin Jil.	66	59	10.7
Las Lomas	22	16	27.3
Pacoc	10	10	-0-
Patzun	107	89	16.9
Sta. Maria Cauque	25	22	12.0
San Marcos P.R.	15	8	46.7
Solola	75	61	18.7
San Lucas T.	59	56	5.1
Cerro de Oro	31	25	19.4
Zaragoza	78	66	15.4
Barberena	50	43	14.0
Cuilapa	77	12	6.5
El Progreso	79	68	14.0
San Lucas L.L.	30	27	10.0
TOTALS	1,472	1,250	15.1

sampling was used in Phase II. A total attrition rate of 15.1 percent occurred in the two years between Phase I and Phase III.

Field Operations

The best conceived research and sampling designs are only as good as they are well-executed. In this investigation every effort was made to adhere strictly to the design and to control the quality of data obtained. Administratively, the following organizational structure was used to manage field work:

1. Senior Resident Researcher - responsible for overall field management, budgets, administration and basic logistics.
2. Field Supervisor - responsible for day-to-day supervision of all field activities, quality control of data and field logistics.
3. Mapping Supervisor/Assistant Field Supervisor/Data Management and Control Supervisor - responsible at various phases during the operation for the above noted areas.
4. Interviewers (8) - responsible for interviewing, coding, cleaning and re-checking data.

In addition, the Co-Principal Investigators spent time in supervising and reviewing data as it was obtained. Further, especially during the training and early stages of data collection, data collection supervisors from INCAP's Division of Human Development were called on for assistance. These two individuals had over 21 years of data collection experience in Guatemala between them. As a consequence of their assistance, interviewers were well-trained.

Nevertheless, there are always problems in data collection, no matter what preliminary cautions are taken and the degree of supervision