

BIBLIOGRAPHIC DATA SHEET

1. CONTROL NUMBER

2. SUBJECT CLASSIFICATION (695)

PN-AAK-203

PA00-0000-6356

3. TITLE AND SUBTITLE (240)

Report on a consultation to develop an evaluation survey to validate the 1981 population and housing census in Guatemala.

4. PERSONAL AUTHORS (100)

Goodman, Roe

5. CORPORATE AUTHORS (101)

Am. Public Health Assn.

6. DOCUMENT DATE (110)

1981

7. NUMBER OF PAGES (120)

27p.

8. ARC NUMBER (170)

GIT 001.422 G653

9. REFERENCE ORGANIZATION (130)

APHA

10. SUPPLEMENTARY NOTES (500)

11. ABSTRACT (950)

12. DESCRIPTORS (920)

Evaluation
Guatemala
Population research
Populations
Surveys

Census
Housing
Bengal

13. PROJECT NUMBER (150)

936590000

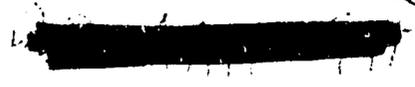
14. CONTRACT NO.(140)

AID/D6PE-C-0053

15. CONTRACT TYPE (140)

16. TYPE OF DOCUMENT (160)

956-5800 - (1)



PN-AAK-203



AMERICAN PUBLIC HEALTH ASSOCIATION
International Health Programs
1015 Fifteenth Street, N.W.
Washington, D.C. 20005

REPORT ON A CONSULTATION TO DEVELOP
AN EVALUATION SURVEY TO VALIDATE
THE 1981 POPULATION AND HOUSING CENSUS
IN GUATEMALA

A Report Prepared By:
ROE GOODMAN

During The Period:
MAY-JUNE 1981

Supported By The:
U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT
(ADSS) AID/DSPE-C-0053

AUTHORIZATION:
Ltr. AID/DS/POP: 7/16/81
Assgn. No. 582098

ACKNOWLEDGMENTS

Of foremost importance to the consultant were the materials supplied by John C. Chao, Office of Population, AID/Washington. The statistical training documentation, called "Popstan" (ISP-TR-40), a case study for the 1980 worldwide censuses of population and housing, was particularly useful. Portions of this documentation (in Spanish) were already in the hands of technical staff at the Guatemalan Statistical Directorate when the consultant arrived.

While working in Guatemala, the consultant received the full support of Director General Martín Carranza Orellana, Dirección General de Estadística, Ministerio de Economía. The consultant gratefully acknowledges the advice and assistance of the population officer, Scott Edmonds, USAID/Guatemala, and the valuable cooperation of the United Nations advisers, Adrian Cartín and Gastón Ormeno, Dirección General de Estadística.

C O N T E N T S

	<u>Page</u>
ACKNOWLEDGMENTS	i
EXECUTIVE SUMMARY	iii
I. INTRODUCTION AND BACKGROUND	
Purpose of the Assignment	1
Itinerary	1
Background	i
II. OBSERVATIONS AND FINDINGS	
Constraints on the Survey Design	3
Size of Sample	4
Selection of the Sample	5
Use of the Supplementary Questionnaire	6
Coding and Analysis of Results	8
Conditioning for Precision Operations	9
III. CONCLUSIONS AND RECOMMENDATIONS	11
APPENDICES	
Appendix A: List of Contacts	
Appendix B: Sampling Methodology	
Appendix C: Instructions for Supplementary Questionnaire, FC-02a	

EXECUTIVE SUMMARY

The purpose of the assignment, which lasted from mid-May to mid-June 1981, was to help design a post-enumeration survey to validate the 1981 Population and Housing Census for Guatemala. Because it was necessary to proceed with the survey as rapidly as possible, subsampling could not be undertaken and the areas which had already been identified for the sample had to be used. The tentative plan was to reuse the questionnaires from the census. This was done, but a suitable supplementary form was designed to facilitate the subsequent matching of survey and census questionnaires in the office.

By early June, a sample of approximately 200 areas, called sectors, had been selected and technical staff were preparing the necessary documentation for budgetary approval. In addition, staff were finalizing the format for the supplementary questionnaire and the instructions for its use. Administrative action to organize field operations had begun, pending the formal approval of the undertaking and the release of the necessary funds. The time which would be required for interviewing in the field would depend on weather conditions; the rainy season had begun when the consultants arrived. It was hoped that most of the field activities could be completed in June.

The procedures outlined in "A Case Study for the 1980 Censuses of Population and Housing," known as "Popstan" (Statistical Training Document ISP-TR-40, Bureau of the Census) were followed closely, ensuring the successful outcome of the Guatemalan evaluation survey. Portions of the most relevant parts of Popstan were translated by the consultant for immediate use; Spanish versions of other parts were already available. Later in the year the Census Bureau made available the remaining documentation in Spanish.

There was not time to subsample the sectors which were chosen. Consequently, there was a pronounced lack of efficiency in the sample because the sampling units which were used were several times larger than they should have been. The original plan was to use a sample approximately half again the size of the first sample. The proposed budget was, therefore, more than adequate. It was recognized that the large number of households which were to be visited would in no way jeopardize the success of the survey and that in subsequent activities at the office (e.g., matching and processing) subsamples could be selected within each sample sector and used if the volume of work was too great to handle efficiently.

The method for processing the survey data was similar to that used to process the census data, with one exception: The data had to be matched in the survey--a major task. A decision was made to omit from the survey the questions on economic activity. This meant that the principal task in coding census data--classifying occupations--also could be omitted from the survey.

The Popstan documentation provided a wealth of information, not only for the matching operation, but also for processing and estimation. With the information it was possible to make adjustments for certain kinds of data missing from the completed questionnaires.

Although the Popstan documents were excellent, providing a "recipe" for completion of the survey, they did not necessarily facilitate progress. In Guatemala, progress in the use of exacting procedures came slowly. The Guatemalans had no experience with such a detailed plan and, somewhat overwhelmed by the extent and precision of the operations, they were not at first confident that the survey could be administered. In fact, it is going to take longer than anticipated to complete the study.

The mixed reactions of the newer technical staff toward the survey are understandable. The leader of the group and the coordinator of the census, Mr. Isaacs, recognized and has stated firmly that a consultant must be present for one week when the matching operation gets under way. The consultant was impressed with the way the team in Guatemala worked to process the census data and design the evaluation survey. It probably will be necessary to make a short visit in late July or early August, and again when estimations are being made.

It is not known whether the evaluation survey is being administered, but, given the leadership of the team in Guatemala, the consultant is certain that the survey can be completed.

I. INTRODUCTION AND BACKGROUND

I. INTRODUCTION AND BACKGROUND

Purpose of the Assignment

As stated in the AID Task Assignment, the purpose of the technical consultation was to provide expert advice in designing a post-enumeration survey to validate the 1981 national census. The scope of work entailed:

- the provision of expert advice to design a sample frame;
- the development of field instructions; and
- the coding and analysis of results.

Early in the assignment, the consultant found it necessary to extend the work on sampling to include the selection of the sample for the post-enumeration (evaluation) survey. He also found it necessary to design a supplemental questionnaire in the format recommended in Part D of Popstan. The original census questionnaire was to be reused in the survey. The supplement was needed to obtain additional information to match the survey questionnaire with the census questionnaire. The format of the supplement had to be integrated closely with that of a matching form in accordance with the highly practical procedures given in Part D of Popstan.

Itinerary

The consultant arrived on May 16. It was intended that he would remain in Guatemala approximately one month. In his initial discussion with Mr. Edmonds, however, the question of the need for a second visit at a later date was raised. The U.N. census adviser, Mr. Cartín, was present at the first meeting. It was agreed that it was feasible to curtail the stay in Guatemala by several days; departure from the country was scheduled for June 9. Much of this report was written later in June.

Background

The Guatemalan Census of Population and Housing identified persons and their habitual places of residence as of March 22, 1981. It was learned that the size of the Guatemalan population was of great importance to the country, in part because a law provided for the addition of new

seats to the national legislature as the population increased. It was necessary to determine the approximate magnitude of errors in the census count, errors which, it was presumed, would constitute underenumeration. Censuses had been taken in 1964 and 1973. The first attempt to evaluate the population census using a post-census survey was made in 1964 and subsequently published. At that time, there were found to be few omissions in the census.

According to a document published by the Guatemalan Statistical Directorate, "Población de Guatemala, Junio de 1980," the estimated population of the country as of June 30, 1980, was 7,262,419; the 1973 census, unadjusted for any omissions, showed a population of 5,160,221.

II. OBSERVATIONS AND FINDINGS

II. OBSERVATIONS AND FINDINGS

Constraints on the Survey Design

When the consultant arrived at the Guatemalan Statistical Directorate, he found that the post-census evaluation survey would have to be organized under certain strict constraints. The forthcoming 1982 elections were one problem. The consultant recognized that with preparations for the elections under way, it would be too risky to wait until late September or October to begin the field interviews. Another problem, and a direct result of the first, was the timing of field operations. These activities had to begin as soon as possible because the rainy season was starting and conditions in the field would only get worse as summer advanced.

Tentative decisions were made to deal with these constraints. The most important was the decision to reuse the 1981 census questionnaire to eliminate the need for time to prepare a new questionnaire for the survey. Money could be saved also by using the old questionnaire because an ample supply of forms was available.

Another tentative decision was to use census enumeration areas, called sectors, as units of sampling. This plan had to be accepted, despite a pronounced inefficiency in sampling and the loss of independence from the census. Ample provision had been made in the budget which allowed for the re-enumeration in the evaluation of up to 3 percent of the census sectors.

Early on, the consultant emphasized the need for a one-page supplementary questionnaire to facilitate the crucial matching operation which was to begin immediately after fieldwork ended. In reviewing the documentation known as Popstan, the consultant found a plan for matching that could be done without transcription of any kind, using a suitable format for one page of the questionnaire.

Because of time constraints, the sampling was inefficient. It was not possible to subsample selected sectors; the sectors had to be re-enumerated in their entirety once the sample was chosen. The sectors contained an average of 150 households; there were more than 150 households in the sectors in many rural parts of the country. The approach was flexible, however, for subsamples could be selected within each sample sector at the matching stage if the volume to be processed was found to be excessively large.

It was, the consultant found, possible to vary the sampling procedure for the municipio (county) of Guatemala (Guatemala City). Separate work in defining area sampling units (also called sectors) had been done in 1979-1980 for an income-and-expenditure survey. The second frame consisted of

sectors which, on the average, were less than half the size of the census sectors. The consultant and his colleagues found that it was feasible to choose the sample for the evaluation survey from this second frame and thus eliminate much of the inefficiency in sampling Guatemala City. Another advantage of the procedure was that it provided a sample which was completely independent of the census enumeration.

For the income-and-expenditure survey, subsampling had been conducted in the new, smaller sectors, but the method which was used was not suitable for the evaluation survey. All the smaller sectors had to be units of sampling for Guatemala City; the units would, perhaps, be twice the size of optimum sampling units, but efficiency of sampling would improve dramatically nonetheless. It was calculated that the population of Guatemala City for which sampling would improve constituted approximately 12 percent of the population of the entire country.

Size of Sample

Part F of Popstan was particularly useful in determining the size of the sample. The documentation revealed that the appropriate method was to select approximately 50 sampling units in each domain of study, each of which would contain approximately 20 (or slightly more) dwelling places. It was decided that an effort would be made to select 50 sampling units for each domain of study, although sampling units which were larger than the optimum size would be used (this fact was disregarded). With this design, extensive subsampling within sectors could be done during processing; the sample size for each domain would remain satisfactory.

Although the evaluation was to provide data for use at the national level, it was found to be feasible to define several domains of study and then to plan for the suitable size of the sample for the different domains. Accordingly, an agreement was reached to have four geographical domains. These consisted of the western and mountainous part of Guatemala, the Altiplano; the northeastern part, or Nor Oriente; the southern part, the Sur; and the major city, Guatemala City. Other domains were identified later after cases from all geographic domains in the two categories urbano (urban) and rural (rural) were accumulated.

It was desirable to use uniform sampling fractions for all domains; calculations showed that a sampling fraction of 1/50 would serve very well. Only for the Sur did this rate produce a sample on the low side; therefore, it was decided that the fraction 1/40 would be used for this domain.

Using the sampling procedures described below, the sampling fractions were applied and 202 sampling units were selected. The number of units selected, by domain, were:

Domain Altiplano	64	Domain Sur	40
Domain Nor Oriente	48	Domain Guatemala	50

Selection of the Sample

Well verified sampling frames (lists of sectors) were on file; thus, the decision was made to use the frames for the survey sample. Provision was made to accommodate minor irregularities in the frames.

Because of difficulties in the field, 10 of the country's 327 municipios had to be excluded from the population (or universe) represented in the evaluation survey. These 10 municipios contained nearly 3 percent of the country's population at the time of the 1973 census. The enumeration sectors of the census for these municipios were completely omitted during the selection of the sample.

For the sectors which had been enumeration areas in the census (i.e., for the entire country, except Guatemala City), the official lists contained preliminary data on the number of households which had been enumerated, and the households were identified as being urban or rural. Some of the census sectors were part urban and part rural, so the categories "urban" and "rural" were subdivided further into "urban," "urban-mixed," "rural," and "rural-mixed." These subdivisions were, however, ignored in the sampling process. The urban sectors appeared first in the list of sectors for a given municipio, and were followed by the rural sectors.

After the list of sectors for a given domain had been prepared, the complete lists were reviewed for irregularities (e.g., notations that a given sector was divided during the enumeration into two sectors or that it contained an especially large number of enumerated households). In a few instances, sectors had been divided, the division was accepted, and another sector was added to the sampling frame. Cases of "very large" or "large" sectors were merely noted for later handling (it was necessary first to determine the number of such cases and was more convenient to make provision for them later on).

It was urgent that the samples be chosen for each of the domains so that necessary administrative and budgetary preparations could be made for field operations. A system requiring one or two days of work for each domain was devised. A recordkeeping system also was needed that would enable technical staff of the Statistical Directorate to follow activities and to have on hand the information they would need for subsequent stages in the survey.

In the absence of a person trained in sampling, the consultant proceeded with the preparations for the sample selections in the Altiplano, Nor Oriente, and Sur. He was joined by one or more members of the survey commission during actual sample selection, which proceeded domain by domain.

Where sectors were unusually large, last-minute steps were taken before the sample was selected to ensure that these sectors would not be over-represented in the sample. In a few cases, the large sectors were subdivided and one part only was included in the sample.

For the municipio of Guatemala, the sectors had been defined for the 1980 survey. It was thought that the sample for the evaluation survey should be a subsample of those sectors which had been chosen in the first stage of the earlier survey. For this purpose a special procedure was required because the sampling had been done with probabilities proportionate to the estimated number of dwelling units in each sector and sampling fractions had varied slightly from zone to zone within the city. A list was prepared of the 352 sectors which had been chosen earlier and new probabilities were calculated that conformed with overall probabilities of 1/50 for Guatemala City. The probabilities were cumulated before the systematic sampling began.

Zones 24 and 25 of the municipio of Guatemala had been excluded from the population sampled for the income-and-expenditure survey. For these two zones, the census sectors were added to the new list. Each sector was given a suitable probability. Selection was made with the assistance of Mr. Lara. Exactly 50 units were chosen by systematic sampling.

Detailed descriptions of the activities were prepared in Spanish for subsequent use by the commission. Sampling worksheets ("sampling guides") also were made a permanent part of the records. (See Appendix B for a translation of these descriptions.)

Use of the Supplementary Questionnaire

The reference date for the survey was supposed to have been the current date. This date, of course, changed each day as the interviewing progressed. The supplementary questionnaire was used to determine which persons had been "non-movers," "in-movers" since the date of the census, or "out-movers." Special efforts were made to find the names of and obtain details on out-movers. If an entire family had moved, it usually was necessary to go to the home of a relative or acquaintance to obtain the needed information. For out-movers, a specific question had to be included in the supplement to tell the enumerator where he might go to obtain reliable information on each person in a household that had been relocated.

No attempt was made in the survey to locate the addresses of in-movers. These persons merely were identified as "in-movers." Past experience has shown that, especially in developing countries, attempts to acquire addresses of in-movers have not been successful.

The supplementary questionnaire included a properly spaced set of lines on which to list all members of a family currently residing at a given dwelling place and another set of lines on which to list out-movers. Out-movers were listed on the reverse side of the form. These lines coincided with the lines on a matching form later designed and used in the office. An illustration is found in Popstan.

A column was required on the questionnaire under which were listed the matching codes recorded on the matching form. By using the census questionnaire, which had been pre-coded, and the supplement, it was possible to process the survey data efficiently.

It was planned that the number of matched out-movers reported in the survey would be used to calculate the percentage of omissions in the census for movers generally, and the results would be applied to totals and categories of in-movers in the estimation of omissions.

Instructions were needed for handling out-movers. It was emphasized that, where dwellings were vacant, the enumerator had to ask particularly about the former residents, when they had lived in the dwelling, and whether or not they were living there at the time of the census.

Experiences elsewhere have shown that more people tend to be missed in evaluation surveys than are missed in a census; in particular, more out-movers than in-movers are missed. Evaluation surveys can, however, provide important data on census omissions. In the current study, which used a dual system and a recommended method of estimation, it was possible to obtain meaningful estimates of omissions, despite the imperfections.

Members of the Commission for the Evaluation Survey recognized that, for political reasons, it might not be feasible to re-enumerate several of the sample sectors. It was agreed that in such cases other sectors would not be substituted and that the records would be used to determine objectively which areas could not be surveyed under existing conditions. Enumerators were instructed to begin their work in the more difficult parts of their sectors and to work intensively in each section so that interviewing would not have to be spread among widely scattered households at the end of the study. It was decided also that if work in a given sector had to be terminated prematurely, an effort would be made to use the portions of the questionnaires completed by that time. (See Appendix C for additional information on use of the questionnaire.)

Coding and Analysis of Results

The method for coding and inputting the survey data resembled closely the method used to process census information. For the survey, however, an additional task was required: matching survey and census questionnaires.

Chapter D-7 in the Popstan document contained information on the procedure for matching. The consultant began to translate this chapter, section by section, but was able to turn over to the commission only about one-tenth of the chapter before he departed. It was not urgent that the chapter be translated. The task is being continued, however, pending receipt of a translation from the Census Bureau. Other parts of the document will be mailed from time to time to the census coordinator of the Statistical Directorate to maintain communication. The coordinator has stated firmly that a consultant must be on site for at least one week when the matching operation begins.

There is little need for coding because the questions on economic activity which were included in the census have been omitted from the evaluation survey. These were the only questions for which detailed coding was required.

The analysis of survey data will differ from the analysis of census data. For the survey it will be necessary to estimate the percentage of omissions in the census. Ample time will be needed to become familiar with the estimation procedures, which are forbidding to many persons. The procedures are covered and explained thoroughly in Chapter D-9 of Popstan.

A translation was made of the section in which the formulas are listed and additional explanations were prepared in Spanish for those who could find time for study.

To estimate the total population of Guatemala at the time of the census, calculations were made for each of the geographic domains. The value X (census total of persons enumerated in each domain) represented excluded persons who were counted even though their names were not known. Persons in the survey were excluded if they could not be matched with persons in the census with no names.

The formulas were used to calculate totals. The goal was to estimate rates of omission, that is, relative, and not absolute, quantities. In the preliminary calculations, the level of the estimated totals (which will be divided one by the other) will be ignored to simplify the work. For calculations within a domain, the value of t_j always can be taken as 1 if the level is disregarded. The value of the estimate of N_1 will be the number of non-movers within the domain. Similarly, the estimates of

N_2 and N_3 and M_1 and M_3 can be simplified by counting only the number of matched persons.

The sample weight, t_j , can be used to calculate estimates at the same level for all domains and to improve the precision of results based on different kinds of imputations and adjustments. At the beginning, it would be useful to use a weight of 50 for t_j for all the geographic domains, except the Sur, for which the weight should be 40. (The sampling fraction in the Sur was $1/40$; elsewhere, $1/50$.)

The formulas in Chapter D-9 are based on the assumption that cases of overenumeration are rare, so rare in fact that they can be ignored completely. In this study, a person in the census questionnaire who was not in the corresponding survey questionnaire (neither non-mover nor out-mover) was counted as an omitted person in the survey. It is necessary to recognize that, depending on the rules which were adopted for the matching, persons may or may not have been "found" in the survey.

The symbols used for the formulas on pages 104 and 105 in Part D, Chapter 9, differ in part from those in Part A, Chapter 13 (available at this time in Spanish). In the latter, the possibility of overenumeration, as well as underenumeration, is considered. Part D should be used because the procedures apply also to the estimation of characteristics of the population. The methodology is not limited only to the estimation of coverage in the census.

Conditioning for Precision Operations

The staff selected for the survey apparently were chosen well, but their modest training did not prepare them to perform the kind of precise operations that were required. Their lack of confidence was not surprising and in fact is characteristic of nationals elsewhere who lack training or experience in the administration of survey questionnaires.

The attitude of the technical staff was good, though they did not seem to be enthusiastic about undertaking survey work that was so precise. Survey work takes time. In a country such as Guatemala, there is no guarantee that the group responsible for processing the census and the evaluation survey will remain on the job in the event of a change in the political leadership. Two persons in the group in Guatemala had worked in the field during the 1964 census, but it is not known whether they were employed by the Directorate in the intervening years.

The brevity of the training for the staff in Guatemala was a constraint, according to one person. The person implied that it was unrealistic to expect a person to absorb a considerable amount of new material in two weeks or less.

The Popstan materials were comprehensive and valuable in two respects. They revealed the need for precision operations and provided an overview of the entire process, step by step. Estimation and variance formulas were included in the materials, but these formulas appeared to be forbidding to staff with little mathematical training and few prospects for permanent employment.

A rigorously devised plan was needed to ensure the success of the matching operation in a reasonable period of time. Certain activities had to be emphasized over others. The work on sampling was kept in perspective, and the staff were content to finish that part of the preparations promptly. A knowledge of the distribution of sample areas by municipio was required to calculate a budget. For this reason, it was also necessary to complete the sampling.

Considerable time was required to design the supplementary questionnaire. The final drafting of the supplement and of the instructions for its use was left in the hands of the Commission for the Evaluation Survey. This group can be depended upon to complete this task.

III. CONCLUSIONS AND RECOMMENDATIONS

III. CONCLUSIONS AND RECOMMENDATIONS

The evaluation survey is for the Guatemalan Statistical Directorate, a major undertaking with considerable promise. In addition to providing objective evidence of, for example, omissions in the census, the survey will provide data on births and deaths in the interim between the census and the survey. The data may be used for demographic and other analyses, as Ferdinand Rath, adviser of the International Labour Organization of the United Nations, has pointed out.

The evaluation survey is a statistical survey. During its administration, many employees in the Statistical Directorate will receive practical training. The technical staff will be justified in taking pride in their work if they successfully complete the survey.

It is recommended that a positive response be given to additional requests from the Statistical Directorate for further assistance. The need has already been expressed verbally for a consultant who could be available for one week or ten days to assist with matching. This activity can and should be started before the consultant's arrival. Additional help may be needed later to follow the imputation procedures for processing the census. However, this and other activities may not be completed on schedule. Problems with timing should be anticipated.

APPENDICES

Appendix A
LIST OF CONTACTS

Appendix A
LIST OF CONTACTS

Guatemalan Statistical Directorate

Dr. Martín Carranza Orellana, Director General
Mario Alfredo Isaacs, Census Coordinator
Ernesto Vargas, Census Processing Group
Antonio Aguirre, Census Processing Group
Aníbal Lara, Commission for Evaluation Survey
Edgar Orellana, Commission for Evaluation Survey
Manuel Pinto, Commission for Evaluation Survey
Josefina Antillón de Vasquez, Technical Chief
Marco Octavio Muralles, Assistant Statistician*

Others

Scott Edmonds, Population Officer, USAID
Adrian Cartín, Census Adviser, United Nations
Gastón Ormeno, Sampling Adviser (Income and Expenditure Survey),
United Nations
Ferdinand Rath, Statistical Adviser to Planning Commission,
International Labour Organization, United Nations

* Candidate for position of sampling statistician.

Appendix B
SAMPLING METHODOLOGY

THE DESCRIPTION AND EXPLANATION THAT FOLLOW ARE TRANSLATIONS OF THE VERSIONS DRAFTED IN SPANISH FOR THE COMMISSION FOR EVALUATION SURVEY, GUATFMALAN STATISTICAL DIRECTORATE. THE WRITERS DID NOT AT ANY TIME PRETEND TO BE EXPERIENCED IN SAMPLING; CONSEQUENTLY, HIGHLY TECHNICAL PHRASING IS AVOIDED, AND THE TEXT IS AS READABLE AS POSSIBLE. THE TEXT ALSO HAS BEEN EDITED FOR CLARITY. A DEFINITION OF SYSTEMATIC SAMPLING, THE BASIC METHOD USED IN THE SURVEY, IS INCLUDED.

Appendix B

SAMPLING METHODOLOGY

Systematic sampling was used in the evaluation survey. This approach is described in the first paragraph of Chapter 8 in the book "Techniques of Sampling," by William G. Cochran.

This method of sampling at first view is seen to be quite different from simple random sampling. Suppose that N units in the population are numbered from 1 to N in some order. In order to select a sample of n units, we take one unit at random from the first k units and from there onward, each k th unit. For example, if k is 15 and the first unit chosen is number 13, the subsequent units will be numbers 28, 43, 58, etc. The selection of the first unit determines the entire sample. This type is called a systematic sample of each k th unit.

A procedure similar to that described above was used in the survey. It is important to note the phrase "are numbered from 1 to N in some order." There is no restriction of any kind on the ordering of the units; they may be arranged in any order desired before systematic sampling is undertaken.

The procedures that will be used to select samples for the domain Nor Oriente and for the City of Guatemala appear to be complex because they reflect the methods for ordering the units and the desire to avoid having to prepare long lists of sampling units (i.e., sectors).

It is feasible to prepare complete lists of the sectors for the purpose of systematic sampling, but the list for Nor Oriente would contain 2,387 units and that for Guatemala City would contain approximately 2,450 units.

It was thought that the method used here would clarify the ordering of the units before systematic sampling. Five to eight pages of details on each domain are included.

The first step in determining the order of the units (sectors) for systematic sampling is to determine the sequence of groups of units, zones for Guatemala City, and municipios for the remaining domains. Often, maps are used in this kind of work; a sequence of the groups is noted by observing the geographic configuration of the groups on the map. In the survey, the enumerators began in Nor Oriente, the domain, in Municipio Puerto Barrios in Departamento Izabal; they then continued with the other municipios--Morales, Los Amates, El Estor, Livingston (in that order), and

proceeded to Departamento Peten, Municipio San Luis, then Poptun, Dolores, Melchor de Mencos, and so forth, following a serpentine route on the map, each municipio being contiguous to that which preceded it on the list. The sequence of municipios for Nor Oriente ended with Municipio Chinautla in Departamento Guatemala.

The same system for ordering municipios was used in two other domains, Sur and Altiplano, and for the zones within the city, Domain Guatemala.

In ordering the sampling units (sectors), a tentative decision was made to accept the existing sequence of sectors within each municipio. It was found to be advantageous, however, to examine what would happen if the sampling were undertaken without modifying the order of the sectors in the municipios. It was found that, except for Guatemala City, which is virtually urban throughout, the proportion of urban to rural sectors in a given sample would not necessarily be the same as that in another sample; that is, if repeated selections of samples were considered, the urban to rural proportions would vary from one sample to another. It would be difficult to control the proportions if the sequencing of the sectors found in the lists of sectors for each municipio were maintained strictly.

To remedy this problem, a procedure called controlled selection (see Cochran, page 175) can be used. The application of this method is so laborious that any sampling statistician hesitates to use it unless a computer program developed for the purpose also is available. There are numerous ways to use controlled selection. In the survey, a less difficult procedure was found which will resolve the problem of proper urban and rural representation. The ordering of the sectors was modified to the degree necessary to ensure uniformity in the urban and rural proportions in systematic samples. This solution was feasible, no matter where the random sampling began.

"Sampling guides" were prepared which specify the sequence of sectors within the sequence of municipios. Use of the system in Nor Oriente is an example. There, it was decided that a fraction of 0.02 of the sectors (i.e., one of each 50 sectors within the sequence of sectors) would be selected. It was convenient to list first the municipios in the order determined, as well as the number of sectors (urban, urban-mixed, rural-mixed, and rural), municipio by municipio, and then to specify successive sets of 50 sectors each (often considered to be minor strata).

For Nor Oriente, the first sets were:

<u>Departamento</u>	<u>Municipio</u>	<u>Number of Sectors</u>
Izabal	Puerto Barrios	(first) 50
	Puerto Barrios	24 (remainder)
	Morales	(first) 26
	Morales	50
	Morales	13 (remainder)
	Los Amates	(first) 37

The first 26 sectors in Morales were used to complete the second set of 50 sectors. To date, the composition of those sectors has not been clarified. If the existing sequence within the municipio were maintained, the first five sectors would be urban and the remaining 21 would be rural, that is, sectors within sections (secciones), as the lists show.

To facilitate the selection of the systematic sample, the sectors in each set are assigned numbers which run from 1 to 50. In this system, the 26 sectors are presumed to have numbers which run from 25 to 50; according to the existing sequence, the numbers for the urban sectors would run from 25 to 29, inclusive. With this system it is not necessary to write the numbers 1 to 50 at the side of each sector which forms a part of each set of 50 sectors.

The sequence of the sectors can be changed within the municipio simply by specifying order numbers which refer to the urban and rural sectors. In general, for the survey, it was necessary only to specify the sequence numbers within the sets of 50 sectors for the urban sectors because the sequence for the rural sectors followed automatically.

In the original order, the first 21 rural sectors in Morales carried the numbers 30 to 50, inclusive (the next 50 rural sectors had the numbers 1 to 13, being in still another set of 50 sectors). Another column in the "sampling guide" was used to specify the order numbers for urban sectors, which always were understood to be within the urban (and rural) sectors. The original sequence was maintained unless the sequence of one or more sectors was changed (a change was noted in the guide).

A detailed analysis was made, and it was found convenient to change the order numbers for the urban sectors in Municipio Morales; the numbers 25 to 29 were abandoned in favor of the numbers 36 to 40. The result was

that the first 21 rural sectors had the numbers 25 to 35 and 41 to 50, both inclusive, instead of the numbers 30 to 50.

It is difficult to explain the method of determining the convenient changes of order numbers. Given the changes introduced, it may be sufficient to show only that good results can be obtained.

Within combined urban areas of the domain (and within combined rural areas), each number from 1 to 50 should appear with equal frequency to ensure correct urban and rural proportions for the sample, assuming no errors have been made. By using the "sampling guide" for Nor Oriente, sequences can be found for the urban sectors, as the following example shows (this is a partial sequence; the codes are for departamentos and municipios).

18-01 Puerto Barrios, 1 - 42

18-03 El Estor, 43 - 50

18-03 El Estor, 1 - 2

16-01 Coban, 3 - 19

15-03 Rabinal, 20

17-03 San Benito, 21 - 34

17-01 Flores, 35-45

16-01 Coban, 46 - 50

19-01 Zacapa, 1 - 12

20-01 Chicuumula, 13 - 41

21-01 Jalapa, 42 - 50

21-01 Jalapa, 1 - 6

15-03 Rabinal, 7 - 14

02-03 San Ag. Ac., 15 - 24

15-03 Rabinal, 25 - 28

02-01 El Progreso, 29 - 38

Etc.

There are various ways to prove that each number from 1 to 50 appears, in its turn, for urban sectors, assuming that no error has been made. A computer program has been developed to order the sampling units, or at least to provide selection patterns on which to base the sequencing. In Nor Oriente it was demonstrated that the numbers entered in the "sampling guide" give a correct solution, which applies equally to rural and urban sectors.

In ordering the sectors, it was feasible to consider especially large sectors (large in terms of numbers of housing units enumerated, sector by sector, in the census). An extreme case was observed in Municipio San Juan Chamelco (Seccion 07, Sector 2), where 766 households were enumerated in the census. By using the method described above, the samplers found that they did not have to include this entire, enormous sector in the evaluation survey. Nor did they have to include an excessive number of other large sectors in the sample.

Where especially large sectors are involved, it is best to assign additional chances (additional sequence numbers) to the sector to ensure that only a part of the sector will be included in the sample survey. If more than one large sector is involved, a method of sequencing can be used to ensure the selection of only the correct proportion of sectors for the sample.

In Nor Oriente, 29 extreme cases were identified. These were sectors with more than 200 dwelling places enumerated in urban areas and more than 250 dwelling places enumerated in rural areas.

The files on the evaluation survey include the "sampling guides" which contain the precise details on the selection of all the sectors for the samples in Nor Oriente, Sur, and Altiplano.

For the sample of Guatemala City, it was advantageous to use other sampling units, and not the census sectors, because sampling materials from the survey of income and expenditures were available for that city. Area sampling units which were less than one-half the size of the census sectors were devised. The smaller sampling units are much more efficient, and their use facilitates the evaluation survey. For the evaluation survey, it was considered advantageous to use a subsample of the units selected in the first stage of the sampling for the earlier survey.

For the survey of income and expenditures, preliminary estimates were made of the numbers of dwelling places in each sample. The probabilities of selection of each unit varied in proportion to the estimated numbers of dwelling places. For the new survey, it was considered desirable to have a sample of units with equal probabilities, that is, each unit would have a precise probability of 0.02.

A new probability was calculated for each of the selected sampling units by dividing 0.02 by the original probability for each unit. A

simplified procedure was found to carry through the calculations of the 352 units selected in the first stage for the municipio of Guatemala. The cumulative sums of the probabilities calculated for the selection of the subsample for the new survey were obtained by systematic sampling. Forty-nine units were selected; an additional unit (part of a census sector) was selected to represent zones 24 and 25, which had not been represented in the survey of income and expenditures. A detailed description of this work is found in the files on the evaluation survey.

The use of smaller units resulted in improved efficiency, as the example of Guatemala City shows. It is anticipated that 2,500 households will be enumerated in the evaluation survey in this city; the sample for Dominio Sur is expected to include approximately 5,000 households--twice the number of households enumerated in Guatemala City. The sample for Guatemala City is based on 50 units, whereas that for Dominio Sur is based on only 40. The data for Guatemala City are expected to be more precise than those for Dominio Sur, but the cost of interviewing in the latter will be far greater because twice the number of households will be surveyed.

Appendix C

INSTRUCTIONS FOR SUPPLEMENTARY QUESTIONNAIRE, FC-02a

Appendix C

INSTRUCTIONS FOR SUPPLEMENTARY QUESTIONNAIRE, FC-02a

Most of the instructions for the enumeration and for interviewing each sample sector are found in the Manual del Empadronador. Additional instructions are given below to obtain answers for Form FC-02a, Supplementary Questions.

One general change always must be kept in mind. The reference date for the enumeration is the date when each dwelling place was visited, and not the date (March 22, 1981) on the printed questionnaire. Several questions have been reworded and are listed below.

- VI. A. Cuantas personas durmieron en este hogar anoche?
- VI. B. Cuantas personas residen habitualmente en este hogar ahora?
- VI. C. Cuantas de estas personas se encuentran en otro municipio o en otro pais anoche?
- VII. 1. Cual es el nombre y apellido de cada una de las personas que residen habitualmente en este hogar ahora?

One other change has been made and is particularly important to remember. The answer under desocupado (vacant) in II. 5 does not mean that the questionnaire has been completed for that dwelling place. One must turn to Form FC-02a, Questions VIII. 3 and 4, to determine whether anyone who was living in the dwelling place on March 22, 1981, has died (Question 3) or moved away (Question 4).

The only questions that refer to the date March 22, 1981 are listed on Form FC-02a, Supplementary Questions. Normally, one would begin with Question 1. The ordering is designed to facilitate the questioning. The purpose of Question 1 and Question 2 is to identify persons who are called non-movers, persons who at the time of the census lived in the same dwelling places as they do now. The individual is asked first if he or she has lived "in this dwelling place" all year long (Question 1). If the answer is negative, the person is asked whether he or she moved there (1) before March 22, (2) on March 22 (3) or after March 22, or (4) whether he or she was born after March 22.

The purpose of the next two questions is to determine whether anyone who was living "here" on March 22 has since died (Question 3) or moved

away (Question 4). With the answer to the latter question, the enumerator will know where to go to obtain the name of (and other essential information on) each person who was living "here" on March 22 but who since has moved away.

The last part of Question 4 will provide limited information on where to go to obtain names and other information. The enumerator subsequently will have to visit the persons and places on which he or she obtains information to answer Question 5 (on the reverse side of the questionnaire).

The questions are similar to those in the census questionnaire, and the same instructions that are listed in the Manual del Empadronador should be followed.