ECUADOR
1994 Sample Design and Proposed 1995 Design

Attachment C

Levinson met with Xavier Dávalos, Director of Socio-Demographic Statistics, INEC; Jorge Eguiguren, Economist, Survey of Living Conditions (SLC); John Mora, Consultant (SLC); Carlos Becerra, International Consultant (SLC); Remigio Burbano, Chief of the Department of Statistical Analysis and Sample Design; and Bolivar Galán, Statistician (SLC), INEC. During this preliminary meeting, general aspects of the SLC were discussed, specifically the sample design of the 1994 SLC.

The main objective of the meeting was to agree on key issues leading to an improved SLC sample design for future rounds of the survey.

Before detailing the decisions taken at the meeting, it is necessary to present a brief description of the 1994 SLC.

A. Background of the 1994 SLC

The *Servicio Ecuatoriano de Capacitación (SECAP* - Ecuadorian Training Agency) signed an agreement with the World Bank to plan and implement the 1994 SLC. The objectives of the 1994 SLC were:

1. Provide the Ecuadorian National Government, national and international institutions, and researchers in general, information on the standard of living of the Ecuadorian population.

2. Provide information to evaluate the impact of the application of governmental social policies on the standard of living of the population.

The target universe of the 1994 SLC is the urban and rural population of Ecuador, that is to say, the population of Ecuador that resides in private housing units, excluding the part of the population that lives in group quarters, such as prisons, convents, military installations, etc. The unit of selection is the housing unit and the units of observation and analysis are the households and the persons in them.

The sample frame for the sample selection is made up of listings obtained from the 1990 Census of Population and Housing and from a subsequent exhaustive update of the frame due to jurisdictional changes which took place in September of 1993.

Six domains of estimation have been defined, all with urban and rural coverage. The definition of an urban area extends to any jurisdiction, known in Spanish as *Cabeceras Cantonales* (CC - Canton Seat) or *Cabeceras Parroquiales* (CP - Parish Seat) with 5,000 or more inhabitants in 1994. Everything that is not included in this definition of urban is considered to be of a rural nature. That is to say, the definition of rural includes:

1. the CCs and CPs with less than 5,000 inhabitants.

2. the periphery of all the CC's and of all the provincial capitals (PC).

3. the dispersed areas of all the CP's.

Taking this definition into account, the sample frame is composed of 116 populated centers, among the provincial capitals, the Canton seats and the Parish seats (rural parishes).

The definition of the six domains of estimation has been drafted taking into consideration the objectives of the survey, that is, including in each domain a group of cities with similar socioeconomic characteristics, following the guidelines established by Carlos Larrea for the populated centers of the country and applying, at the same time, certain modifications to it to meet the needs of the 1994 SLC.

The domains of estimation which have been defined are:

1. **Quito**

   Self-Representing City.

2. **Guayaquil**

   Self-Representing City.

   These two cities have been considered as separate domains of estimation because they contain 48 percent of the urban population of the country and have socioeconomic characteristics that are unique and different from the ones in the rest of the country.

3. **This domain contains groups 2 and 3 of Larrea, that is, all the provincial capitals in the Sierra (Mountain Area) and Costa (Coastal Area) regions, except Guaranda, but including Santo Domingo.**

   This domain contains the following cities:

Sierra		Costa	
Cuenca	Ambato	Machala	Santo Domingo
Riobamba	Loja	Portoviejo	Esmeraldas
Ibarra	Tulcán	Esmeraldas	Babahoyo
Latacunga	Azogues		

4. **This domain contains the Larrea groups 4 and 5.**

Sierra		Costa	
Otavalo	Sangolqui	Manta	Milagro
Cayambe	Atuntaqui	Quevedo	Salinas
Gualaceo	Machachi	Santa Elena	Naranjito
Guano	Cotacachi	Montecristi	Chone
Conocoto	Catamayo	Jipijapa	Lago Agrio
Cumbaya	Pomasqui	Huaquillas	Anconcito
Tumbaco			

This domain contains the Larrea groups 6 and 7, plus the city of Guaranda, that Larrea places in group 8, which is mostly made up of cities located in the eastern part of Ecuador (*Oriente*).

Sierra	Costa	
SAN GABRIEL	PASAJE	ARENILLAS
CAÑAR	DAULE	PIÑAS
PIMAMPIRO	BALZAR	ZARUMA
BAÑOS	EMPALME	MARC.MARIDUEÑA
CATACOCHA	VENTANAS	PORTOVELO
PELILEO	QUININDE	DURAN
PILLARO	SAN LORENZO	LA LIBERTAD
SAN MIGUEL	NARANJAL	PLAYAS
GUARANDA	CALCETA	EL CARMEN
PUJILI	ROCAFUERTE	EL TRIUNFO
CALDERON	SANTA ANA	BUENA FE
LA TRONCAL	EL SALITRE	PEDRO CARBO
SAN ANTONIO	PAJAN	POSORJA
MACARA	PUEBLO VIEJO	EL GUABO
LA MANA	MUISNE	YAGUA. NUEVO
SALCEDO	SANTA ROSA	PASCUALES
CARIAMANGA	VINCES	SAMBORONDON
ALAUSI	BAHIA DE CAR.	J.L.TAMAYO
GUARANDA	LOMAS DE SARG.	MONTALVO
	SAN VICENTE	JARAMIJO
	TOSAGUA	SANT LUCIA
	PALESTINA	VALENCIA
	MOCACHE	PUERTO LOPEZ
	A.B.MORENO (JUJAN)	BALAO
	PEDERNALES	CHARAPOTO
	RICAURTE	

6. This domain is made up of cities in the last Larrea group. These are small towns with great state dependency.

Oriente

Puyo	Francisco de Orellana
Tena	Yantzaza
Macas	Zamora
Shushufindi	Sucua

The rural part of the domains is made up of housing units from:

1. the periphery of the CC's with 5,000 or more inhabitants in 1994.

2. Cantones with less than 5,000 inhabitants in 1994.

3. Dispersed enumeration areas of the CP's included in the urban domains.

Considering that the rural population of the peripheries of the cities and cantones of the urban domains and the population in the rural parishes with 5,000 or less inhabitants in 1994 is intimately linked to the government seat (cabecera) to which it belongs (that is, it is assumed that the above-mentioned rural population is part of a rural domain dependent upon the urban domain), the rural sample was defined as associated to the urban domains. This means that the selection of housing units was mainly carried out from rural enumeration areas that are within the jurisdiction of urban domains (periphery and dispersed areas) of the selected cities.

Thus, the rural part that corresponds to each one of the domains of study defined in the urban areas is made up as follows:

Domain 1

1. Rural part of Quito
2. Pedro Moncayo
3. Pedro Vicente Maldonado
4. San Miguel de los Bancos

Domain 2

1. Rural part of Guayaquil

Domain 3

I. Sierra

The following list corresponds to Cantones with CC's of less than 5,000 inhabitants in 1994.

Girón	Nabón	Pucará
Mira	Espejo	San Miguel Urcuquí
Saraguro	Saquisilí	Cevallos
San Fernando	Chambo	Mocha
Santa Isabel	Colta	Patate
Oña	Guamote	Quero
Paute	Pallatanga	Tisaleo
Biblián	Penipe	Chinchipe

II. Costa

Atacames	Baba	Sucre

Domain 4

I. Sierra

Sig Sig

II. Costa

Chillanes	Cascales	Putumayo
Simón Bolívar	Pangua	Cumandá
Gonz. Pizarro	Flavio Alfaro	Sucumbíos

Domain 5

I. Sierra

Chaguarpamba	Espindola	Gonzanamá
Chimbo	Caluma	El Tambo
Bolívar	Chunchi	Celica
Puyango	Sozoranga	Zapotillo
Pindal	Quilanga	

II. Costa

Echeandía	Atahualpa	Balsas
Chilla	Marcabelí	Las Lajas
Eloy Alfara	Colimes	Palenque
Catarama	Junín	Pichincha

Domain 6

Oriente

Limón Indanza	Palora	Santiago
Huamboya	Gualaquiza	Aguarico
Archidona	El Chaco	Joya de los Sach.
Quijos	Mera	Santa Clara
Nangaritza	Yacuambi	El Pangui

For the 1994 SLC a two-stage design was used for the self-representing cities and a three-stage design was used for the rest of the urban domains. For the rural part, a three-stage design was used.

In the urban area, the Primary Sampling Unit (PSU) is the city which was selected with probability proportional to size measured in terms of housing units.

For Domains 1 and 2 (Quito and Guayaquil), the probability of selection of each city is equal to 1. Before selecting enumeration areas in these two cities, a socioeconomic stratification was conducted in order to improve the efficiency of the design.

Domains 3, 4 and 5 are divided into natural regions (Sierra and Costa). These natural regions were used to assign sample proportional to the number of housing units in each natural region.

Domain 6 is exclusively made up of cities in the eastern (*Oriente*) part of the country, although Nueva Loja, a city located in the *Costa* of Domain 4, has

been excluded because it presents characteristics similar to the other cities in the domain.

The enumeration areas were chosen with equal probability, that is, independently from the number of housing units in the selection stratum.

The ultimate sampling units (USU) are the housing units in all the domains.

A workload was defined as the task to be undertaken by a team of 3 interviewers during a predetermined period of time. To that effect, the workload was defined as the task to complete in 6 enumeration areas. The number of housing units selected from every urban enumeration area for interview was 9 and in the rural area 12. That is, the required task is to interview 54 housing units in 6 enumeration areas in 28 days and in two rounds. In the rural area, the number of housing units grows to 72 in 6 enumeration areas (EAs), but the period of time is reduced to 24 days in two rounds. The reason the interview was done in two rounds can easily be explained by the sheer size of the questionnaire and the fear of obtaining a high non-response rate if a complete interview was attempted in one round.

The workloads were used to adjust the sample size in each domain and as a measure of size to determine the probabilities of selection of the cities in the domains.

Table 1 below provides the number of EAs and the number of housing units in sample by domain of estimation and by urban/rural EA.

Table 1

Distribution of the Urban and Rural Sample Enumeration Areas and of Housing Units by Domain of Estimation

Domain	Urban			Rural		
	Locality	Sector	No. HUs	Locality	Sector	No. HUs
I.	Quito	84	756	Qiuto Periferico	6	72
				S.M. Los Bancos	6	72
				Pto. Quito	6	72
	TOTAL I URBAN	84	756	TOTAL I RURAL	18	216
II.	Guayaquil	84	756	Gyquil.Periferico	1	12
				Chongon	11	132
				Gomez Rendon	6	72
	TOTAL II URBAN	84	756	TOTAL II RURAL	18	216
III.	III. SIERRA.			III SIERRA		
	Ibarra	6	54	Huachi Grande	1	12
	Ambato	6	54	Sta. Rosa	3	36
	Riobamba	6	54	Pilahuin	2	24
	Cuenca	12	108	Valle	2	24
	Loja	6	54	Baños	2	24
				Sinincay	2	24
	III. COAST			III. COAST		
	Esmeraldas	6	54	Sto. Domingo	3	36
	Santo Domingo	6	54	Alluriquin	3	36
	Machala	6	54			
	Portoviejo	6	54			
	TOTAL III URBAN	60	540	TOTAL III RURAL	18	216
IV.	IV SIERRA			IV. SIERRA		
	Cayambe	6	54	Cayambe Periferico	3	36
	Sangolqui	6	54	Olmedo	2	24
	Catamayo	6	54	Cangahua	4	48
	IV. COAST			IV. COAST		
	Manta	6	54	Flavio Alfaro	9	108
	Milagro	6	54			
	Quevedo	6	54			
	Santa Elena	6	54			
	Montecristi	6	54			
	TOTAL IV URBAN	48	432	TOTAL IV RURAL	18	216
V	V. SIERRA			V. SIERRA		
	La Troncal	6	54	Baños Periferico	2	24
	San Antonio	6	54	Ulba	2	24
	Baños	6	54	Rio Negro	2	24
	V. COAST			V. COAST		
	Ventanas	6	54	Santa Lucia Perif.	12	144
	Playas	6	54			
	Santa Rosa	6	54			
	San Vicente	6	54			
	Santa Lucia	6	54			
	TOTAL V URBAN	48	432	TOTAL V RURAL	18	216
VI	ORIENTE			ORIENTE		
	Puyo	12	108	Tena Periferico	8	96
	Tena	6	54	Puerto Napo	4	48
	Macas	6	54	Sev. Don Bosco	3	36
	Zamora	6	54			
	Sucua	6	54	General Proaño	3	36
	TOTAL VI URBAN	36	324	TOTAL VI RURAL	18	216
TOTAL	TOTAL URBAN	360	3240	TOTAL RURAL	108	1296

The probability of selection of a housing unit (HU) in an EA was obtained by multiplying the probability of selection of a city (P_c) by the probability of selection of an EA given that the city was selected ($P_{s/c}$) by the probability of selection of a HU given that the EA was in sample ($P_{v/s}$). If P_v denotes the probability of selection of the HU, then:

$$P_v = (P_c) * (P_{s|c}) * (P_{v|s})$$

($P_{s/c}$) and ($P_{v/s}$) are conditional probabilities, that is, ($P_{s/c}$), for example, is the probability of EA s being in sample, given that the city c is in sample.

The EAs were selected with equal probability and ($P_{v/s}$), in the urban area, was defined as:

$$P_{v|s} = Pv_{ijr} = \frac{12}{V_{ijr}}$$

The weighting factor f is equal to the inverse of the probability of selection P_v, that is,

$$f = \frac{1}{P_v}$$

B. Sample Design of the 1994 SLC

During the aforementioned meeting several issues were raised regarding the sample design of the 1994 SLC. In the following paragraphs, these issues are fully described and recommendations are made, if warranted.

Levinson reviewed the 1994 SLC sample design and made the following recommendations:

1. In the 1994 SLC a socioeconomic stratification was used in Quito and Guayaquil to improve the efficiency of the sample design. To this effect, the sampler utilized the "General Index of Availability of Basic Services by Enumeration Area". However, the availability of basic services is quite widespread and, therefore, it is not a good discriminatory tool of economic condition.

   RECOMMENDATION: Study the socioeconomic stratification used in the 1994 Income and Expenditure Survey, since it is known to have produced good results. The above-mentioned stratification produced two socioeconomic strata (high and low) and it seems to have been satisfactory for the objectives of the survey.

2. Even though the sample design of the urban part of the 1994 SLC is solid, one should not choose enumeration areas with equal probability because if the EAs differ greatly in size, the estimates can suffer from a big bias.

RECOMMENDATIONS REGARDING SELECTION PROCEDURES: Choose the EAs for the 1995 SLC with probability proportional to the size of the EA. In order to do this, follow the steps outlined below:

a. Within each domain of estimation (or study) and socioeconomic stratum (if any) order the EAs geographically in a serpentine manner to obtain an implicit stratification within the selection stratum.

b. Register the total number of HUs for each EA, M_{hi}, from the 1990 Census of Population and Housing, where h represents the socioeconomic stratum and i is the EA. This number will serve as measure of size for the selection of the EAs.

c. Cumulate the measure of size through the socioeconomic stratum (if any). The last cumulated number, (M_h), will be equal to the number of HUs in the socioeconomic stratum h of the domain of estimation under consideration.

d. Obtain the number of EAs, (n_h), to be selected from the corresponding socioeconomic stratum (if any).

e. Determine the interval of selection, (I_h), by dividing the cumulative number of HUs in the socioeconomic stratum (M_h) by the number of EAs to be selected.

$$I_h = \frac{M_h}{n_h}$$

f. By means of a random number generator, obtain a random number (A_h) between 1 and the interval of selection (I_h), inclusive. It is necessary to keep two decimal places if the interval of selection is not a whole number. Many random number generators produce random numbers between 0 and 1. If that is the case, multiply the random number obtained by I_h in order to obtain A_h.

g. Determine the sample EAs by using the following formula:

$$A_h + (k-1) \times I_h$$

$$k = 1, \ldots, n_h$$

rounding up.

The kth selected EA will be the one which has the cumulative measure closer to this quantity, without going over it.

h. The probability of selection of the EA is given by:

$$P(EA) = n_h * \frac{M_{hi}}{M_h}$$

In addition, the conditional probability of an HU given that the EA is in sample is given by:

$$P(V|S) = \frac{m_{hi}}{M'_{hi}}$$

The probability P that an HU is in sample is given by:

$$P = P(EA) * P(V|S) = n_h * \frac{M_{hi}}{M_h} * \frac{m_{hi}}{M'_{hi}}$$

where:

m_{hi} = number of housing units to be chosen in the i-th EA.

M'_{hi} = number of housing units in the EA that are occupied permanently obtained from the updating operation.

The basic weighting factor, w_{hij}, is the inverse of the probability of selection. That is,

$$w_{hij} = \frac{1}{P}$$

3. In the 1994 SLC, the probability of an urban HU was defined as:

$$Pv_{ijr} = \frac{12}{V_{ijr}}$$

even though only 9 HUs were selected per EA. Further up in the report, the following formula is introduced:

$$Pv = \frac{12}{V} * \frac{11}{12} = \frac{11}{V}$$

where it is indicated that "if 11 HUs were used, the initial probability 12/V of the total number of HUs (V) in the EA must be corrected with the factor 11/12, after which the probability would be equal to 11 divided by the total number of HUs in the EA (V)." One can infer from the phrase in quotes that of the 9 HUs to be interviewed, 2 had to be replaced using the replacement HUs previously selected to that effect and that, in the end, only 9 interviews were obtained inside the EA in question.

RECOMMENDATION: Under the previous premise, the probability of selection of the HU inside the sample sector is given by:

$$Pv = \frac{9}{V}$$

since only 9 HUs were interviewed, and not 11.

If it were to happen that, after having used up all the replacement HUs, less than 9 interviews are obtained, the basic weight will have to be adjusted by the corresponding non-response factor.

The use of the basic weight only assumes the existence of 9 completed interviews in the urban EAs and 12 in the rural EAs. In most cases that will be true, since there are reserved or replacement units available to be used in case of noninterviews in the original selected HUs. However, in extreme cases, all the reserved units may be utilized before completing the desired number of interviews. If this were to happen, it will be necessary to adjust the basic weight of the responding units to compensate for the loss of units in the sample. This adjustment will take place at the EA level, where a certain level of socioeconomic homogeneity among HUs can be assumed.

Therefore, the nonresponse adjustment factor (for the urban area) will be:

$$F_{1hi} = \frac{9}{r_{hi}}$$

where:

F_{1hi} = adjustment to the basic weight for all the interviewed HUs in the i-th EA of the h stratum.

r_{hi} = number of interviewed HUs, including replacements, in the i-th EA.

$r_{hi} \leq 9$

This procedure assumes that the responding HUs and the nonresponding HUs do not, on the average, differ significantly with respect to socioeconomic characteristics and, in addition, that the nonresponse rate within the EA is low. The more we deviate from these premises, the higher the unquantified amount of bias will be. To control this bias, a restriction must be imposed: if the factor F_{1hi} is higher than 2.00, that is, if the response rate in the EA does not reach 50 percent, the adjustment must be made at the socioeconomic stratum level (or stratum used to select the EAs), putting together all the EAs in the substratum to carry out the calculation. Since there are no substrata at the rural level, the adjustment has to take place at the level of *parroquia rural*.

4. In the 1994 SLC other adjustments to the basic weight are not contemplated, such as those due to changes in the occupancy status of the HU.

Since the selection of the HU is based upon a list of occupied HUs which is prepared about 1 or 2 months before the interviews, the occupancy status of some HUs could change during the interim. Or maybe some of the HUs were classified wrong during the listing. Every time the one-to-one correspondence between the selected HU (defined in the listing) and the interviewed HU does not exist, the basic weights will have to be adjusted to account for these discrepancies. In addition, since the field work requires, a constant workload of 9 interviews per EA (in the urban area), there may be adjustments to the basic weight.

At the time of the interviews, five situations related to the occupancy status of the HU may arise:

- the HU is occupied by a new household;
- the HU is occupied but the members of the household are temporarily absent;
- the HU becomes ineligible, that is, vacant, seasonal, destroyed, or nonresidential;
- the HU is divided into more than one household;
- there is a consolidation of two or more HUs.

- In the case of a new household, the interviewer would simply complete the questionnaire for the new household. There would be no need for a weight adjustment since the probability of selection of that household would not be affected.

- In the case of a household that is absent during the interview period, the interviewer would choose a reserve unit and complete a questionnaire for it. No adjustments to the initial basic weight are necessary.

However, if the interviewer cannot get an interview from the reserve unit, then the nonresponse adjustment factor F_{1hi} has to be applied.

- In the case of an ineligible HU, that is, vacant, seasonal, destroyed, or nonresidential, the interviewer would replace it with a reserve unit in order to maintain a constant workload. But since in theory an ineligible unit should not be in sample (and should not be part of the frame in the first place), it will be necessary to deflate the initial basic weight in order not to overestimate the universe. Consequently, the basic weights of all the households in the EA will have to be multiplied by the following correction factor:

$$F_{2hi} = \frac{9 - b_{hi}}{9}$$

where,

F_{2hi} = adjustment to the basic weight for all the interviewed households in the i-th EA.

b_{hi} = number of ineligible sample units in the ith EA.

- In the case of a divided HU, that is, a unit associated with only one household at the time of listing which, at the time of interviewing corresponds to more than one household, the interviewer would select one of the households at random for interviewing in order to maintain a constant work load. As a result, the true probability of selection of that household would be smaller than presumed. Therefore, the basic weight for that particular household should be multiplied by the following adjustment factor to reflect the true probability of selection:

$$F_{3hij} = m'_{hij}$$

where,

F_{3hij} = adjustment factor applied individually to the j-th divided HU in the i-th sample EA.

m'_{hij} = number of households found in the j-th HU at interview time.

This factor is not used if all households found in the HU are interviewed, as is the case in the 1994 SLC and subsequent rounds of this survey.

- In the case of consolidation, there are three possibilities. To illustrate, let's assume that within a housing structure two housing units, HU-01 and HU-02, are listed and that HU-01 is selected in sample. At the time of the interview, the interviewer realizes that HU-01 and HU-02 have been consolidated into one housing unit, that is, HU-01 and HU-02 are occupied by just one household.

  If the household associated with HU-01 in the listing is the one now occupying the consolidated space (that is, HU-02 left the dwelling unit), the interviewer would proceed with the interview. No adjustments would be necessary in this case.

  On the other hand, if the household listed in association with HU-02 is the one now occupying the consolidated space (that is, HU-01 left the dwelling unit), the interviewer would have to classify HU-01 as no longer existing; that is, ineligible, and would replace it by one of the reserve units. In this case, the adjustment factor, F_{2hj}, mentioned before would have to be applied to the basic weight of all the interviewed households in the EA.

  However, if a third (new) household moves into the consolidated space of HU-01 and HU-02 (that is, both HH-01 and HH-02 have left the consolidated space), the interviewer would have to proceed with the interview since he would not be able to distinguish the first case from the second. The weight for the interviewed household, in this case, would have to be multiplied by 1/2 to correct for the fact that its probability of selection was doubled. The general adjustment factor needed in this case, which would be applied individually to the affected household, is:

$$F_{4hij} = \frac{1}{m'_{dhij}}$$

  where,

  F_{4hij} = weight adjustment factor applicable to the j-th interviewed household of the i-th sample EA, in the case of a consolidated space with indistinguishable predecessor units.

  m''_{hij} = number of HUs listed separately for the space currently consolidated at interview time.

  In the case of the 1995 SLC all the households found in the sample housing units will be interviewed. Therefore, this factor will not be applied.

5. In the 1994 SLC, the basic weight was adjusted by an "increment factor", denoted by D. This factor was created because "it was noticed in the

field that there exists an increment of housing units inside the EAs and outside as well, which would contribute to the increment in the cities. However, even though the intrasector growth is taken into consideration when selecting HUs from the number of HUs obtained during the cartographic update, the extrasector growth is not considered. In certain areas of certain cities, this extrasector growth is higher than the intrasector one". However, it must be noticed that Ecuador is completely divided into EAs and, therefore, it is not necessary to apply this factor, since the basic weight contemplates this type of situation.

RECOMMENDATION: Do not use this increment factor.

6. Since some of the EAs have a very high number of housing units, it may happen that some EAs may fall in sample more than once. Since big EAs cannot be segmented in the office, if any large EA were to fall in sample it will have to be divided into segments of about 150 HUs each in order to control the enumerator workload and the variances. The procedures to be followed to segment a big sample EA are detailed in the next section.

7. Segmentation of Large Sample EAs

A quick count of the number of HUs in the sample EAs will take place before the listing operation. This operation is called "quick count" because the enumerator counts, in a systematic way, the number of HUs in the EA in order to get an idea of the size of the EA in terms of number of housing units. After the quick count, the enumerator prepares a listing of all the households in each EA in order to determine the eligible households from where the sample will be ultimately selected. In the 1995 SLC, the new sample EAs will be listed in order to update the frame. However, some EAs which were in sample last year will be kept for 1995 but will not be relisted in order to save time and resources. Nevertheless, the further away in time from the census date, the more the EAs are susceptible to change and, consequently, do not reflect the reality at the time of the survey.

Since there are EAs in the frame which have at least 200 HUs, it's very possible that some of these EAs may fall in sample more than once, as mentioned before. If this were the case, the enumerator will perform a quick count of the HUs in the sample EA in order to segment the EA according to the guidelines shown in Figure 1. To carry out the segmentation of a sample EA, following the steps outlined below:

i. Identify in the field each sample EA with 200 or more HUs.

ii. Make a quick count of the number of HUs in the sample EA using very detailed maps of the EA, in order to be able to segment the EA. It is extremely important that the segments created from the sample EA have fixed and identifiable boundaries.

Figure 1
Segmentation of Sample Enumeration Areas

Number of Housing Units from Quick Count	Number of Segments to be Created
200-299	2
300-449	3
450-599	4
600-749	5
750-899	6
900 or more	7

Again, it is imperative that each segment have clear and identifiable boundaries in the field so that the interviewer can locate the segment and the HUs within it. Now, if an EA were to have, say, 300 HUs, it is not strictly necessary to try to obtain two segments of 150 HUs each if the boundaries of the segments are not clear and identifiable. For instance, one of the segments could have 120 HUs and the other 180, as long as the boundaries of both segments are clear and identifiable in the maps.

iii. Assign a sequential number from 1 to k, where k represents the number of segments to be obtained from the sample EA. If the sample EA is in sample twice, segment the EA and select two segments at random. If the sample EA is in sample 3 times, segment the EA and choose three segments at random, and so forth.

iv. It may happen that a sample EA that has 200 HUs in the frame shows a larger number of HUs during the quick count operation. If this were the case, the same instructions given for the big sample EAs will have to be followed.

8. Difficult Enumeration Areas

Sometimes it happens that for a given reason (war, floods, or any other unknown or unforseeable reason), an EA cannot be enumerated. In that case, it is necessary to have procedures available to replace an EA by another. If this were necessary, follow the procedures detailed below:

i. Within the domain of estimation and the socioeconomic stratum (if any), select a random number between 1 and the cumulative measure of size, M_h.

ii. Locate the interval within which this random number falls. If that EA is already in sample, reject it.

iii. Repeat step 1 above. Locate the interval within which this random number falls. If the EA was not selected previously, that EA will now be in sample.

iv. All the procedures outlined in this report with respect to selected EAs must now be followed by the new replacement EA. That is, if the number of HUs is very large, the sample EA must be segmented. One or more segments must be chosen at random if the circumstances so require it. The new sample EA must now be listed as was done for the other EAs in the sample.

9. The rural part of the 1994 SLC was defined as associated to the urban domain and, therefore, a rural sample was only obtained in the jurisdictions of the urban domains (peripheries and dispersed areas) of the selected cities. This methodology can create a bias because the assumption is that the housing units of the periphery of the selected cities have socioeconomic characteristics which are similar to the HUs of the peripheries of the cities that were not selected and to the housing units of rural areas not contiguous to the peripheries.

RECOMMENDATION: Choose an independent sample in the rural area, as it is detailed in Section C of this report.

C. Recommendations and Guidelines for the Sample Design of the 1995 Survey of Living Conditions

1. Domains of Estimation

Six domains of estimation were defined in the 1994 SLC, as mentioned in Section A of this report. However, one of the main objectives of the SLC is to analyze the value of the estimates and the factors that determine the values, in order to implement appropriate policies. It is for that reason that it is convenient to create independent domains of estimation in the urban and rural sectors, since these two domains have socioeconomic and demographic characteristics that are intrinsically very different. Therefore, for the 1995 SLC, the same domains of estimation used for the 1994 SLC will be maintained but, in addition, separate estimates will be obtained for the urban and rural sectors of each domain. This explicitly creates 12 domains of estimation, as shown in Figure 2 below.

Figure 2

Domains of Estimation for the 1995 SLC

Domain	Sector	
	Urban	Rural
1	X	X
2	X	X
3	X	X
4	X	X
5	X	X
6	X	X

Of course, from these domains others (aggregations) can be created such as the national total, the urban total, the rural total and the natural regions, *Costa*, *Sierra* and *Oriente*.

The urban frame of the 1994 SLC was composed of populated centers (provincial capitals, canton seats and parish seats) with a minimum of 5,000 people in 1994 (that is, centers with at least 4,650 inhabitants in the 1990 census).

The urban frame of the 1995 SLC is composed of all populated centers with a population of at least 4,550 in the 1990 census, since in 1995, due to demographic growth, these populated centers will have reached the 5,000 population mark. A demographic growth of 2.2 percent per annum was considered to arrive at the above-mentioned figures.

During the 1994 SLC the periphery of the populated centers was considered to be of a rural nature, but when the results were analyzed after the survey, it was verified that the socioeconomic characteristics of the households in the peripheries were very similar to those of their corresponding populated centers.

Therefore, for the 1995 SLC it is recommended to undertake a study of the peripheries of the populated centers in the domains of estimation to determine if they should be part of the urban domain or the rural domain. To that effect, the Data Processing Department of INEC was asked to produce a data base with 1990 census data containing certain specific characteristics to determine the classification of the peripheries of populated centers. The following variables were requested to be included in the data base:

- Percentage of HUs with dirt floors.
- Percentage of HUs that use wood or charcoal as kitchen fuel.

- Highest level of education of head of household.
- Overcrowding (number of people per household and per bedroom).
- Percentage of population engaged mainly in agricultural activities.
- Percentage of the population engaged in industrial, commercial and service-related activities.

Since the Domains Quito and Guayaquil are mostly of an urban nature, it was decided to use the same classification for urban/rural that was used in the 1994 SLC. This will allow the scarce resources to be concentrated in analyzing the peripheries of domains of estimation 3, 4, 5 and 6. Since the degree of ruralization becomes more pronounced as we go from domain 3 to 4 and from 4 to 5 and from 5 to 6, the first domain to be studied carefully should be domain 3. For each city in domain 3, a close scrutiny of the above-mentioned variables should be done in order to classify the periphery as either urban or rural. If the periphery of domain 3 were to be fully classified as urban, then the same scrutiny should be done for domain 4. If the periphery of domain 4 were to be classified as rural, there is no need to continue with domains 5 and 6, because their peripheries will automatically qualify as rural.

It is necessary to mention that when assigning the periphery to a city, the complete periphery must be assigned to either the urban or the rural area. That is, if it is decided that a periphery is mostly urban, then the whole periphery becomes urban for purposes of this survey. Likewise, one should not use the same criteria to assign peripheries in all domains, but rather the assignment of the periphery should be done on a domain basis. For example, the percentages, level of education and overcrowding in domain 4 that allow for the classification of the periphery as urban may be lower than the ones needed to classify the periphery as urban in domain 3.

This assignment is extremely important, since the more homogeneous the domains of estimation, the more efficient the sample design and the lower the variances to be obtained.

2. Sample Frame

The Republic of Ecuador is divided geografically in four natural regions: *Sierra*, *Costa*, *Amazónica (u Oriente)* and *Insular*. For political and administrative purposes, Ecuador is divided into *Provincias* (Provinces). Each *Provincia* is divided into *Cantones*. Each *Cantón* is divided in turn into *Parroquias Urbanas* (Urban Parishes) and *Parroquias Rurales* (Rural Parishes). In addition, each *Cantón* has a *Cabecera Cantonal* (Cantón seat) which, in turn, is composed of one or more *Parroquias Urbanas*. In each one of the *Provincias*, one of the *Cabeceras Cantonales* functions as capital of the *Provincia*. All the *Cabeceras Cantonales* are by definition of an urban nature.

For census and statistical purposes, the *Parroquias Rurales* are divided into Rural Sectors which have an average of about 75 HUs. Likewise, the *Cabeceras Cantonales* are divided into Census Zones which, in turn, are made up of about 10 EAs which contain on the average about 150 HUs each.

The sample frame for the 1995 SLC comes from the 1990 Census of Population and Housing which has, in September 1993, been updated in terms of politico-administrative jurisdictions and of population numbers.

Areas of the country that for whatever reasons should not be part of the sample frame should be identified and eliminated from the sample frame. These areas include border areas under dispute, far-away and difficult-to-enumerate areas which impose an undue burden and cost on the survey operations. Likewise, the data user should be informed about these areas to allow him/her to take into consideration the limitations encountered at the time of analysis and utilization of survey data.

The unit of analysis is the household and the persons in them; the unit of observation is the household. The universe under study consists of all the households in the Republic of Ecuador, excluding the insular region and the areas previously mentioned.

In order to reach the sample HUs, a sample of cities has to be selected and within them a sample of EAs. From the latter, the sample HUs will be selected. Therefore, it is necessary to obtain a sample frame that:

- is up-to-date, complete and exact with respect to coverage of the population under study.

- contains sufficient elements to allow for the identification of the sample housing units, and

- the most important aspect, that gives a positive probability of selection to every housing unit, fundamental requirement of probability sampling.

The maps and the listing must provide a complete count of the HUs within the sample EA. While the maps must provide a visual identification of the boundaries of the sample EAs and of the HUs within them, the listing must contribute the following information:

a. Name of household head.
b. Address of the HU.
c. Brief description of the HU.
d. Occupancy status of the HU.

However, since the 1995 SLC lists housing units and not households, the "Name of Household Head" will be omitted from the listing operation.

The occupancy status of the housing unit is necessary in order to determine the HUs that are occupied permanently, since these HUs are the only valid ones under the coverage requirements. To that effect and to better control coverage errors, the enumerator will have to list all the HUs in the EA and register the occupancy status of each one of the HUs in the sample EA:

a. Occupied permanently.
b. Seasonally occupied and with persons seasonally absent or for

other reasons.
c. Vacant.
d. Under construction.
e. Demolished or being demolished.

The identification of the valid housing units will take place in the office after the listing operation.

3. Stratification

The sample frame will be divided into more or less homogeneous strata to improve the efficiency of the sample design. The first stratification level will be used to provide reliable estimates for each one of the 12 domains of estimation.

It's very important in the big urban centers to stratify further the domains of estimation to improve even more the efficiency of the estimates. To that effect, substrata which are homogeneous with respect to key characteristics can be created. For the 1995 SLC, socioeconomic strata can be created in domains 1 and 2 (Quito and Guayaquil).

A very successful socioeconomic stratification has already been created for Quito and Guayaquil for the 1994 ENIGHU. A similar stratification scheme can be used in this instance, specifically suited to the needs of the 1995 SLC. Further information can be obtained from Remigio Burbano or Consuelo Aguinaga, at INEC.

4. Assignment of EAs to the Domains of Estimation

a. Urban Part and Periphery

Table 1 presents the distribution of the 1994 SLC sample in the urban and rural areas. The sample size in the urban area will be the same in 1995, that is, around 3,240 housing units.

The cartographic update that must take place in each sample EA to select the final sample of housing units is a very expensive operation, since it requires a lot of time and a lot of resources. Therefore, it was decided to keep in sample in 1995 most of the 1994 sample urban EAS.

Since it is very likely that the size of the urban part of certain domains may change if the periphery is added to them, it is necessary to determine the number of EAs that will be kept in the original urban area (that defined for the 1994 SLC) and the number of EAs that will be assigned to the periphery. The allocation of the number of EAs to the original urban part and the urban periphery will be done proportionally to the size of each part, keeping in mind the limitations on the number of workloads, that is, multiples of 6. An example will help determine how to carry out this operation. Let's suppose that n EAs have been originally assigned to the urban area of a domain of size X_1 housing units; let's assume that the periphery has been added to the urban domain and that it has a size X_2 housing units. Then, the number of EAs

that should be kept in the original urban part, denoted by n_1, is equal to:

$$n_1 = n * \frac{X_1}{X_1 + X_2}$$

and the number of EAs that should be allocated to the urban periphery, denoted by n_2, is equal to:

$$n_2 = n * \frac{X_2}{X_1 + X_2}$$

These numbers should be adjusted so that they are exact multiples of 6, since a workload is defined as the work to be carried out in 6 EAs.

The n_1 EAs that will remain in the original urban area will have to be chosen at random from the n original EAs.

Since a socioeconomic stratification will be used in Quito and Guayaquil to classify each EA into a socioeconomic stratum (high, low, medium), the number of EAs allocated to each socioeconomic stratum will also be done proprtionally to the size of the socioeconomic stratum in terms of number of housing units in each stratum. Since it is necessary to preserve the sample EAs chosen for the 1994 SLC, it will be very difficult to allocate each EA to a socioeconomic stratum and, therefore, we may run the risk of losing representativeness in the socioeconomic strata. The EAs corresponding to Quito and Guayaquil have already been classified according to the "General Index of Availability of Basic Services in the EAs", but these classification does not seem to have proven useful to discriminate EAs.

b. Rural Part and Rural Rest

In order to obtain estimates at the rural domain level, the size of the sample in the rural area has been increased from 1,296 HUs to more or less 2,500 HUs, which represent about 210 sample EAs. These 210 EAs were allocated to the domains proportionally to the size of the rural domain. Even though Levinson did not have at the moment the exact sizes of the rural domains, and keeping in mind that the number of EAs assigned to each rural domain must be a multiple of 6, a preliminary allocation was carried out and it is presented below in Figure 3.

Figure 3

Allocation of Sample EAs to the Rural Domains

Rural Domain	Númer of Eas
1	24
2	18
3	54
4	24
5	66
6	24

In domains 3, 4 and 5, the sample EAs allocated must be divided proportionally to the size of their respective natural regions, *Sierra* and *Costa*, always keeping in mind that the number of EAs must be a multiple of 6 in order to maintain the workloads. The rural sample size is then obtained by multiplying the total number of EAs by 12, which yields 2,520 housing units.

5. Selection of Cities

Since it is desired to maintain most of the 1994 SLC EAs in the 1995 SLC, the cities that have been selected in sample for the 1994 SLC will also be in sample for the 1995 SLC.

6. Selection of Sample EAs

a. **Urban EAs**: the sample EAs that will be selected in the peripheries classified as urban must be chosen following the methodology presented in Section B.2. of this report.

b. **Rural EAs**: the selection of the rural EAs will be carried out in three stages. First, we will choose a sample of geographic areas grouped in cantones and composed of the rural area of the cantones that make up the urban domains 3, 4, 5 and 6, plus the cantones whose seats have less than 4,550 inhabitants, excluding from all these areas the cabeceras parroquiales with more than 4,449 inhabitants. Let's call each one of these geographic areas Special Rural Zones (SRZ). The selection of SRZs will be done with probability proportional to the number of work loads.

To maintain a more stringent supervision and to keep costs down when carrying out the updating of rural EAs, it is better to select within each sample SRZ a sample of *parroquias rurales* with

probability proportional to the number of workloads. By using this intermediate step, we gain more control and less dispersion of the sample. Once these sample *parroquias rurales* have been chosen, allocate the number of work loads proportionally to (a) the number of housing units in the *cabeceras parroquiales* (areas arranged in blocks) and (b) to the number of housing units in the dispersed areas. This way we are sure to obtain sample in both, the areas arranged in blocks and the disperse areas of the *parroquias rurales*.

After having chosen the *parroquias*, the next step is the selection of the EAs with probability proportional to the size of the EA (size measured in number of HUs as of the 1990 census date). Within the sample EAs, 16 housing units will be selected with equal probability, and subsequently 12 HUs will be chosen from the original 16, keeping 4 HUs as reserve units.

Now, let

P	=	Probability of choosing a housing unit.
P_i	=	Probability of choosing a Special Rural Zone.
P_j	=	Probability of choosing the *parroquia rural* j given that the Special Rural Zone i is in sample.
P_k	=	Probability of choosing the k-th EA given that the *parroquia rural* j and the Special Rural Zone i are in sample.
P_l	=	Probability of choosing the housing unit l given that the k-th EA and the *parroquia rural* j and the Special Rural Zone i are in sample.

Therefore, the initial probability of selection of a housing unit is given by:

$$P = P_i \, P_j \, P_k \, P_l$$

7. Variance Calculation

For the calculation of variances the use of the IMPS module CENVAR is highly recommended. This module has the advantage of providing outputs which present the estimates, the sampling error, the coefficient of variation, the 95% confidence interval, the design effect and the number of observations used to obtain the sampling error.

8. General Recommendations

a. Precision of Estimates Obtained in the Rural Domains

The rural sample of the 1995 SLC has been designed with a predetermined sample size of 2,520 housing units. Therefore, the sample size has not been gotten, taking into consideration the precision to be obtained for

key estimates. However, the data from the 1994 SLC can provide estimates and sampling errors, even though there were no rural domains of estimation in 1994. Therefore, since we have doubled the rural sample size, we can expect a much lower sampling error for the 1995 rural estimates.

b. Sample of Urban Housing Units

Since the questionnaire to be used during the 1995 SLC is voluminous, it is necessary to select a new sample of housing units within the urban EAs so as not to impose a severe burden on last year's respondents. If we were to return to the same households, we would run the risk of getting a very high non-response rate.

c. Control and Supervision

A household survey is a statistical enterprise of immense proportions and is composed of several implementation stages, such as the questionnaire design, the sample design, the field work, data entry, publication of results, just to name a few. Therefore, the introduction of nonsampling errors in the different stages can severely distort the results, due to the undue accumulation of these nonsampling errors. It is for this reason that it is extremly important and essential for the success of the survey to maintain a very stringent control and a close supervision on each and every phase of the operation to minimize the number and level of the nonsampling error.

d. EA Updating

Updating is recommended for all EAs that have shown significant changes in the number of HUs during the updating operation of 1994. For those EAs, a new listing operation should be undertaken to determine the occupancy status of each housing unit in the EA.

e. Special Comment

Although Levinson has utilized the number of sample housing units in the urban area (9) as an example to present various formulas, it must be kept in mind that this number must be replaced by 12 when we deal with rural EAs. Likewise, in each urban sample EA a sample of 12 housing units should be chosen, keeping 3 housing units as reserved units. In the rural area, the number of sample housing units will be 16, keeping 4 units as reserve. It is also important to mention that the reserve units should be used as a last resort. That is, every possible effort should be made to interview the original sample units, before making use of the reserve ones.

Lastly, Levinson wishes to thank the work group for the common effort and the diligence shown to obtain a solid 1995 SLC sample design during this very short period of work.