

**BASIC INFORMATION DOCUMENT**

**Timor-Leste**

**2007 Timor-Leste Survey of Living Standards**

**Directorate of National Statistics**

**May 2008**

## **Introduction**

The 2007 Timor-Leste Survey of Living Standards (TLSLS) is the second national survey of living standards for Timor-Leste. The first national survey, the Timor-Leste Living Standards Survey (TLSS), was undertaken in 2001 during the months of August to November. The 2001 TLSS had a modest though nationally representative sample of 1800 households from 100 sucos. Being the first national living standards survey of its kind following the independence referendum of August 1999, the TLSS provided a wealth of information on living conditions in the country as an input into the first National Development Plan. The second national living standards survey, the TLSLS, has been undertaken to update this information and is also expected to provide an input into the development of the second National Development Plan. The TLSLS is a comprehensive multi-module survey, which encompasses a very broad scope of topics that typically would be covered under more specialized surveys such as the Demographic and Health Survey, the Multiple Cluster Indicators Survey and a typical labor force survey.

The purpose of the present document is to provide potential data users with the background information they need to understand the cross-sectional component of the TLSLS and how to use it properly. First, information on the sample design used in the survey is provided. Then the fieldwork organization is briefly discussed as well as the estimation of the sampling weights. Finally, a description of the data set and supplementary documentation is offered.

## TLCLS sample design

The TLCLS sample was designed to have two components: (i) a cross-sectional component of 4,500 households selected with the intention of representing the current population of Timor-Leste, and (ii) a panel component of 900 households, where half of the 2001 TLSS sample of 1800 households are randomly selected and re-interviewed, with the purpose of evaluating changes in the living conditions for the same set of households between the two surveys. However, this panel component is not being released at this time, so it will be neither covered in the rest of the documentation nor included in the data files. The cross-sectional component is expected to provide independent estimates for rural and urban areas of each of five recently defined regions, which are groups of districts (see Figure 1) defined as follows:

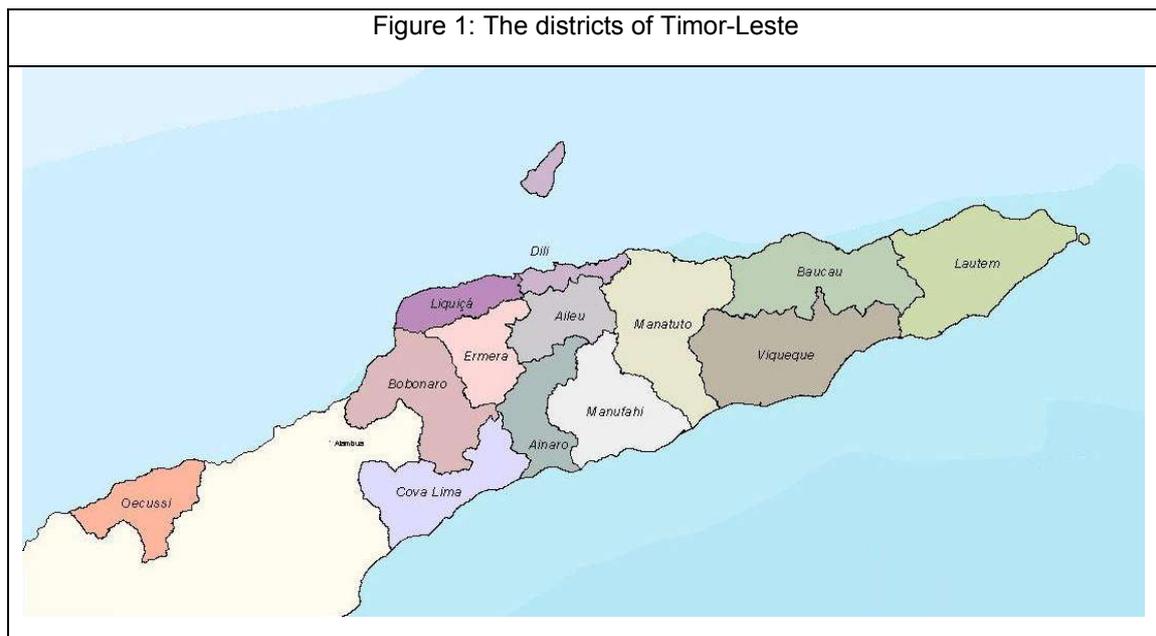
Region 1: Baucau, Lautem and Viqueque;

Region 2: Ainaro, Manufahi and Manatuto;

Region 3: Aileu, Dili and Ermera;

Region 4: Bobonaro, Cova Lima and Liquiçá; and

Region 5: Oecussi.



The cross-sectional sample is selected in two stages. In the first stage, 300 Census Enumeration Areas (EAs) are selected as the primary sampling units (PSUs). In the second stage, 15 households are selected in each EA. The design recognizes ten explicit strata – the Urban and Rural areas in each of the five regions. Table 1 shows the allocation of the 300 cross-sectional PSUs among them.

Table 1: Distribution of enumeration areas for the TLSLS cross-section sample			
	<i>Number of enumeration areas</i>		
	Rural	Urban	Total
<b>Region 1: Baucau, Lautem and Viqueque</b>	35	25	60
<b>Region 2: Ainaro, Manatutao and Manufahi</b>	35	25	60
<b>Region 3: Aileu, Dili and Ermera</b>	35	37	72
<b>Region 4: Bobonaro, Cova Lima and Liquica</b>	35	25	60
<b>Region 5: Oecussi</b>	28	20	48
<b>Timor-Leste</b>	<b>168</b>	<b>132</b>	<b>300</b>

This particular allocation resulted from the following line of reasoning:

- In spite of their different populations and total number of households, sampling theory dictates that a sample of the roughly the same size (60 EAs) should be allocated to each region in order to produce estimates of similar quality for each of them.
- A similar case could have been made for allocating a sample of the same size (30 EAs) to urban and rural areas within each region, but since the definition of urban and rural areas outside Dili was still a matter of discussion, it was decided to opt for an allocation closer to proportional: 25 EAs in Urban areas and 35 EAs to Rural areas.
- Region 5 represents a special case. It is composed of a single district of difficult access (Oecussi, see Figure 1) that ought to be the responsibility of a dedicated team. This imposed a total sample size of 50 EAs for this region, of which only 48 can be allocated to the cross-sectional component since the panel component contains two EAs in Oecussi.
- The capacity thus liberated to visit an additional 12 EAs in the rest of the country was devoted to reinforce the urban sample in Region 3, where Dili is located.

The first sampling stage used the list of 1,163 Census Enumeration Areas (EAs) generated by the 2004 Census as a sample frame. Within each stratum, the allocated number of EAs was selected with probability proportional to size (*pps*) using the number of households reported by the census as a measure of size. No efforts were made to append the smaller EAs to neighboring EAs, or to segment the larger EAs in order to make the size of the primary sampling units (PSUs) more uniform.

The second sampling stage used an exhaustive household listing operation in all selected EAs as its sample frame. Sample households in each EA were selected from the list by systematic equal probability sampling.

As a result of the relatively large sampling fraction in some of the strata, certain large EAs were selected more than once by the *pps* procedure adopted at the first sampling stage. In fact, the cross-sectional sample only consists of only 269 (rather than 300) different EAs. This necessitated selecting a multiple of 15 households (rather than just 15 households) in the EAs that were selected more than once.

The final cross-sectional sample consists of 4,477 households. Table 2 shows the distribution of the total TLSLS sample across the rural and urban areas of the five main regions in the country. The sample can be considered representative at national level as well as at the level of the ten domains represented by the rural and urban areas of the five regions.

	Rural	Urban	Total
Region 1 : Baucau, Lautem, Viqueque	524	375	899
Region 2 : Ainaro, Manatuto and Manufahi	517	374	891
Region 3 : Aileu, Dili and Ermera	522	552	1,074
Region 4 : Bobonaro, Cova Lima and Liquiçá	520	375	895
Region 5 : Oecussi	419	299	718
Timor-Leste	2,502	1,975	4,477

Lastly, it may be helpful to clarify the definition of urban and rural areas. At the time of the 2001 TLSS, 71 of Timor-Leste's 498 sucos were conventionally qualified as urban, of which 31 sucos in the Dili and Baucau districts were qualified as major urban centers. By the time of preparation of the sample design for the 2007 TLSLS, 60 of the 498 sucos defined by the 2001 Suco Survey were conventionally qualified as urban. The partition of the country into sucos was also modified in September 2004. With the amalgamation of several sucos, the original 498 sucos were now collapsed into 442. Many of the rearrangements took place in urban areas with the result that the 60 "old" sucos are now considered urban only constitute 38 "new" sucos (Appendix 1 provides a list of the 60 sucos that are now considered urban).

## Fieldwork

The TLSLS was designed to run over a period of a full year in order to better account for any seasonal variation in different indicators. In addition, the fieldwork was designed to be more or less evenly spread throughout the country over the year. The TLSLS was launched on March 27, 2006. However, after about eight weeks of fieldwork, the survey had to be suspended due to the outbreak of conflict in the country. The survey was resumed on January 9, 2007 and survey operations progressed without interruption since then. Fieldwork for the survey concluded on January 22, 2008.

At the time of the resumption of the survey, a decision was made to revisit the households who were interviewed in 2006 prior to the interruption of the survey. In particular, 351 households had been visited in 2006. Of these, 317 households were revisited during December 2007-January 2008. The remaining 34 households could not be found at the time of the revisits, and instead an additional 41 new households were interviewed as replacement households. In order to maintain a sample for a continuous period of a year, the final TLSLS sample thus excludes the 351 households interviewed in 2006 and instead includes the 358 revisited or replaced households.

Given the challenges of the turbulent political and security situation during some periods in 2007, the fieldwork schedule had on occasion to be slightly modified to accommodate concerns of security and feasibility of fieldwork. Despite this, as seen in Table 3, the distribution of the sample by month of interview and by region and rural and urban areas indicates a sample that is well-spread through the year, which should allay any concerns of intra-year seasonality.

	Region 1: Baucau, Lautem and Viqueque	Region 2: Ainaro, Manatuto and Manufahi	Region 3: Aileu, Dili and Ermera	Region 4: Bobonaro, Cova Lima and Liquiçá	Region 5: Oecussi	Timor-Leste
January 2007	60	90	75	87	58	370
February	91	60	75	90	45	361
March	75	59	105	45	60	344
April	58	45	45	60	45	253
May	75	132	90	135	75	507
June	60	74	105	88	60	387
July	60	74	164	60	60	418
August	45	119	58	60	60	342
September	60	88	90	45	60	343
October	120	30	89	75	76	390
November	105	60	90	45	59	359
December 2007	60	45	45	60	30	240
January 2008	30	15	43	45	30	163
Total	899	891	1,074	895	718	4,477

## Selection probabilities and raising factors

For the cross-sectional sample of TLSLS, the selection probabilities and raising factors are determined in accordance with the sample design described below.

The probability of selecting Census Enumeration Area  $ij$  in stratum  $i$  is

$$p_{ij} = \frac{m_i n_{ij}}{n_i} \quad (1)$$

where  $n_{ij}$  is the number of households in the EA (as reported by the 2004 Census),  $n_i$  is the total number of households in the stratum (also as per the 2004 Census) and  $m_i$  is the number of EAs selected in the stratum.

The probability of selecting household  $ijk$  in EA  $ij$  of stratum  $i$  is

$$p_{ijk} = p_{ij} \frac{15}{n'_{ij}} \quad (2)$$

where  $n'_{ij}$  is the number of households in the EA, as per the household listing operation.

The raising factor or weight  $w_{ijk}$  for household  $ijk$  is the inverse of the selection probability  $p_{ijk}$ . If the number  $n'_{ij}$  of households found at the time of the listing operation were equal to the number  $n_{ij}$  recorded by the census in all EAs, the sample would be self-weighted in each stratum, with a constant raising factor equal to  $n_i/15m_i$ . In practice the numbers  $n_{ij}$  and  $n'_{ij}$  will seldom be equal but often close to each other, meaning that the samples will not be exactly self-weighted, but quite approximately so.<sup>1</sup>

The household weights are further adjusted such that the population totals as estimated from the full sample match the demographic projections for mid-2007 for each stratum. This corresponds to a mid-2007 total population for Timor-Leste of 1,047, 632 persons.<sup>2</sup>

---

<sup>1</sup> Strictly speaking, the above formulae are valid only when the size of the EA is such that it can be selected at most once by the *pps* procedure. However, the artifact of selecting 15 $t$  households in the second stage whenever an EA is selected  $t$  times in the first stage has the effect of making them applicable to compute raising factors even for the large EAs where that may not be the case. Formula (2) may be inadequate if the actual size  $n'_{ij}$  of EA  $ij$  happens to be less than 15. In that (quite unlikely) case, all households in the EA will need to be visited, and  $p_{ijk}$  simplifies to  $p_{ij}$ .

<sup>2</sup> This population total relates to the medium-level projection in DNE (2007), Population Projections 2004-2050: Analysis of Census Results, Report 1, General Population Census of Timor-Leste 2004.

## Description of data files and supplementary documentation

The final database of the cross-sectional component of the household survey is organized in 25 Stata data files (see Table 4 below). The names of the files refer mostly to the section of the questionnaire that is being covered, with the exception of two files: *hhold.dta* and *indiv.dta*. These two data files have already merged all the information available in the survey at the household or individual level. Within each of any data file, the names of the variables reflect the question to which they are referring to, i.e. q08d03 corresponds to section 8, part D, question 3. The unit of observation of the files varies according to the type of information included in the file. For the purposes of merging information from different files, all files include identification variables that allow for a proper matching among the different units of observation.

An additional file called *basicvars.dta* has been included. Its unit of observation is the household and contains a set of variables that identify various geographical areas, different administrative levels and the interview date. It also comprises all the relevant survey design variables such as strata, primary sampling units, sampling weights and household size. In order to produce any estimation, this file must be merged with the raw data file of interest using the unique household identification number, a variable called *hh\_id*.

Table 4: 2007 TLSLS data files

1	<i>hhold.dta</i>	Household level data: Cover, Housing (Section 2), Transfers, borrowing and savings (Section 10 - Q1-13) and Subjective wellbeing (Section 13)
2	<i>indiv.dta</i>	Individual level data: Household Roster (Section 1), Education (Section 5), Health (Section 6), Fertility (Section 7), Employment (Section 8), Aids (Section 14) and Anthropometrics (Section 15)
3	<i>service.dta</i>	Section 2 - Housing, Part C: Ownership and expenditures
4	<i>facility.dta</i>	Section 3: Access to facilities
5	<i>h_provider.dta</i>	Section 6: Health, Part C: Access to health care providers
6	<i>food.dta</i>	Section 4: Consumption & expenditures, Part A: Weekly food consumption
7	<i>non_food.dta</i>	Section 4: Consumption & expenditures, Part B: Monthly/annual non-food expenditure
8	<i>durables.dta</i>	Section 4: Consumption & expenditures, Part C: Durable goods
9	<i>employment.dta</i>	Section 8: Employment, Part A: Jobs
10	<i>enterprise.dta</i>	Section 8: Employment, Part D: Self-employment & business
11	<i>plot.dta</i>	Section 9 - Farming, Part A: Plots
12	<i>crop.dta</i>	Section 9 - Farming, Part B: Crops
13	<i>ag_implement.dta</i>	Section 9 - Farming, Part E: Farming equipment (Q.1)
14	<i>ag_equipment.dta</i>	Section 9 - Farming, Part E: Farming equipment (Q.3-6)

15	ag_input.dta	Section 9 - Farming, Part C: Agricultural inputs
16	ag_labour.dta	Section 9 - Farming, Part F: Labour (Q.2-5)
17	by_product.dta	Section 9 - Farming, Part F: Labour (Q.8-9)
18	forestry.dta	Section 9 - Farming, Part D: Forestry
19	livestock.dta	Section 9 - Farming, Part G: Livestock
20	boat.dta	Section 9 - Farming, Part H: Fishing (Q.3-6)
21	fish_exp.dta	Section 9 - Farming, Part H: Fishing (Q.11)
22	assistance.dta	Section 10 - Transfers, borrowing and savings (Q14-15)
23	savings.dta	Section 10 - Transfers, borrowing and savings (Q16-18)
24	other_income.dta	Section 11: Other income
25	comm_group.dta	Section 12: Social capital
26	basicvars.dta	Key variables at the household level regarding location and survey design.

The supplementary documentation comprises the following four files:

- The questionnaire (*TLALS\_2007\_Questionnaire\_Eng.pdf*),
- The codebook (*TLALS\_2007\_Codes\_Eng.pdf*),
- The interviewer's manual (*TLALS\_2007\_Field\_manual\_Eng.pdf*), and
- An Excel spreadsheet that contains a detailed list of all the variables included in each of the 26 data files (*TLALS\_2007\_DataFilesOrg.xls*).

## **APPENDIX**

Table A.1: Names of the 60 urban sucos in 2006

<b>District: Aileu</b>	<b>Posto: Aileu</b>
010110 ..... Seloi	010113 ..... Hurairaco
(the last two now collapsed into a single suco called Seloi Manere)	
<b>District: Ainaro</b>	<b>Posto: Ainaro</b>
020101 ..... Ainaro	020408 ..... Maubisse
<b>District: Baucau</b>	<b>Posto: Baucau</b>
030201 ..... Tiri Lolo	030208 ..... Caicido
(the last two now collapsed into a single suco called Tiri Lolo)	
030202 ..... Bahu	
<b>District: Bobonaro</b>	<b>Posto: Maliana</b>
040603 ..... Ritabou	040605 ..... Holsa
<b>District: Covalima</b>	<b>Posto: Suai Kota</b>
050502 ..... Laconac	050508 ..... Debos
050509 ..... Vila	
(the last three now collapsed into a single suco called Debos)	
<b>District: Dili</b>	<b>Posto: Cristo Rei</b>
060201 ..... Culuhum	
060202 ..... Centro Benemauk	060204 ..... Becora
060207 ..... Ailok	
(the last three now collapsed into a single suco called Becora)	
060203 ..... Fatuahi	060208 ..... Camea
(the last two now collapsed into a single suco called Camea)	
060205 ..... Hera	060210 ..... Bidau Santana
<b>District: Dili</b>	<b>Posto: Dom Aleixo</b>
060301 ..... Loscabubu	060304 ..... Suleur
060306 ..... Malinamoc	
060310 ..... Rai Naca Doco	
(the last four now collapsed into a single suco called Comoro)	
060303 ..... Nazare	060307 ..... 12 Novembro
060606 ..... Naroman	
060608 ..... Isolado060611	Moris Dame
(the last five now collapsed into a single suco called Bairo Pite)	
060302 ..... Beira Mar	(now called Fatuhada)
060308 ..... 7 Decembro	(now called Kampung Alor)
<b>District: Dili</b>	<b>Posto: Nein Feto</b>
060501 ..... Monumento	(now called Bidau Lecidere)
060507 ..... Talera Hun	(now called Acadiru Hun)
060502 ..... Asucaí Lorosae	060503 ..... Solo
060504 ..... Santa Cruz	
(the last three now collapsed into a single suco called Santa Cruz)	
060506 ..... Inur Fuik	060509 ..... Lahane Oriental
(the last two now collapsed into a single suco called Gricenfor)	
060505 ..... Meira	060508 ..... Bemori
(the last two now collapsed into a single suco called Bemori)	
<b>District: Dili</b>	<b>Posto: Vera Cruz</b>
060604 ..... Mascarinhas	
060605 ..... Rumbia	(now called Caicoli)
060602 ..... Hanso Hatora	060607 ..... Haksolok
(the last two now collapsed into a single suco called Vila Verde)	
060305 ..... 28 Novembro	(now called Colmera)
060309 ..... 20 Maio	(now called Motael)
060601 ..... Alto Hospital	060603 ..... Bairo Alto
(the last two now collapsed into a single suco called Lahane Occidental)	
<b>District: Ermera</b>	<b>Posto: Ermera Kota</b>
070201 ..... Poetete	070206 ..... Talimoro
<b>District: Liquiça</b>	<b>Posto: Liquiça</b>
080201 ..... Dato	
<b>District: Lautem</b>	<b>Posto: Lospalos</b>
090301 ..... Fuiluro	
<b>District: Manufahi</b>	<b>Posto: Same</b>
100301 ..... Letefoho	100302 ..... Babulu

Note: Each suco is identified by a geocode with 2 digits for the district, 2 digits for the posto within the district and 2 digits for the suco within the posto.