

BASIC INFORMATION DOCUMENT

TIMOR LESTE SURVEY LIVING STANDARDS 2 – EXTENSION (TLSLS2-X) 2008

Introduction

In 2007-2008 a multi-topic household survey, the Timor Leste Survey of Living Standards (TLSLS2) was conducted in East Timor with the main objectives of developing a system of poverty monitoring and supporting poverty reduction, and to monitor human development indicators and progress toward the Millennium Development Goals. Information collected in the TLSLS2 questionnaire included: household information, housing, access to facilities, expenditures/consumption, education, health, fertility and maternity history, employment, farming and livestock, transfers, borrowing and saving, other income, social capital, subjective well-being, AIDs and anthropometrics.

The TLSLS2-X extension survey was designed to re-visit one third of the households interviewed under the TLSLS2 2007-08 to explore different facets of household welfare and behavior in the country, while also being able to make use of information collected in the TLSLS2 survey for analytic purposes.

The four new topics investigated in the extension survey are:

Risk and Vulnerability: This section is designed to help us understand the dimensions and sources of household-level vulnerability to uninsured risks in Timor Leste, and the efficacy and welfare effects of various risk-management strategies (prevention, mitigation, coping) and mechanisms (private as well as public, formal as well as informal) households do (or do not) have access to. The work in Timor Leste is part of a program of analytic work and policy dialogue throughout the EAP region, more information on which can be found on the World Bank website.

Land Degradation and Poverty: This section of the questionnaire is designed to identify proximate causes of deforestation through land use patterns and links with poverty; understand strengths and failures of common land resource management institutions (property rights, enforcement); understand the impact of the Siam Weed problem on household welfare.

Justice for Poor: The Justice for the Poor/Access to Justice (J4P/A2J) module of the survey will serve mainly as an initial diagnostic for project development in the country. The topics we would be interested in covering would be Dispute Processing/Resolution; Social Legal Norms and Perceptions of Efficiency in Government (Local, Sub-District, District and National level).

Access to Financial Services: The financial service work has the following two objectives: (i) to collect data on access to and use financial services (savings and credit), both formal and informal, and (ii) assess the quality of information on access to financial services obtained from head of households vs. from all adults – i.e. is there a bias introduced by not asking all household members, do the characteristics of the head or the household affect this (gender, age, nuclear family, urban, education levels, wealth, etc.).

Survey Questionnaires

The household questionnaire was structured as follows:

<u>Book 1</u>	
Section 0	Pre-Printed Cover Page
Section 1	Household Information*
Section 2	Agriculture*
Section 2A	Plots
Section 2B	Land Management
Section 2C	Investments
Section 2D	Forest Use
Section 3	Finance
Section 3B	Individual Financial Information
Section 4	Shocks and Vulnerability
Section 4A	Incidence of Shocks and Household Responses
Section 4B	Future Shocks
Section 4C	Preventive Health
Section 4D	Program Participation
<u>Book 2</u>	
Section 0	Cover Page
Section 3	Finance
Section 3A	Household Financial Information
Section 5	Justice for the Poor
Section 5A	Community Trust and Decision Making
Section 5B	Opinion and Perceptions of the Law
Section 5C	Local Institutions
Section 5D	Dispute Resolution

As one of the objectives of the survey was to measure how much the head of the household knows about the financial activities of all household members, and to be able to produce gender disaggregated analytical work, it was decided that concurrent interviews would be conducted with the head of the household and his/her spouse, to prevent them from contaminating the responses of the other. To accomplish this, the interview was conducted in two parts. During the first part, one interviewer would complete the roster and agricultural modules with the household head, or the household member best able to respond to these questions. They would then schedule a time to return to the household to interview the household head and spouse separately. Two separate books were therefore used.

This survey also included the randomized assignment of the justice and vulnerability modules to the household head or spouse. This was done using a randomized number and designation printed on the cover of the questionnaire. Additionally, this designation was used to assign the household either the short (household head only) or long (all household members over the age of 15) finance module.

The pre-printed random number designated the household into an “odd” or “even” category. The randomization was generally well followed with the exception of one team. All of the interviews conducted by this team are included in the “questionable” category. Any gender disaggregated analysis should therefore exclude the “questionable” households and include the “re-interview” households, using sampling weights w_2 .

* The household information and agriculture sections are based on information from the TLSLS2 survey, with some questions from that questionnaire repeated for new household members and new plots.

All households were asked sections 1 and 2 to the respondent best able to answer. Interviewers asked even households the justice module to the household head, the vulnerability module to the spouse, and the finance module to all household members. Interviewers asked odd households the vulnerability module to the household head, the justice module to the spouse, and the finance module to the household head only.

The division of the questionnaire into household head and spouse was undertaken to guarantee the separation of the two respondents. During field work, while a priority was placed on assigning female interviewers to female respondents and male interviewers to male respondents, this was not always possible due to scheduling conflicts.

To ensure that the correct households were re-surveyed during the TLSLS2-X, the covers of the questionnaires were individually pre-printed with the household location information and the designation of "odd" or "even". Additionally, on the reverse side of the cover, the roster and plot lists from the TLSLS-2 questionnaire were printed as reference.

The roster for the TLSLS2-X listed all current household members. Interviewers compared this information to the pre-printed roster information from the TLSLS2 survey to identify new members. New members were asked a series of questions relating to age, marital status, education, etc, which were skipped for existing members. All members were asked questions related to preventative health care. Interviewers are then asked to compare the new roster with the pre-printed and determine the whereabouts and reason for leaving of all household members in the TLSLS2 survey who were no longer part of the household.

Note: In the case of a female headed household, or a male headed household without a spouse, the procedures changes for the justice and vulnerability modules. The vulnerability module was asked to the most appropriate respondent of the specified gender, while the justice module was skipped. In some cases, a second choice was not available for the vulnerability module and it was also skipped. The actual respondent can be linked to the roster through their personal identification number (pid). Also note that surveys were marked as complete if the interviewer followed the correct procedures, not if all sections are actually completed. Therefore, some surveys marked complete will be missing information for the vulnerability or justice modules.

Sample Design

Sampling

Sampling for the TLSLS2 – Extension survey was a sub-sample of the original TLSLS2 sample.[†] The TLSLS2 field work was divided into 52 “weeks”, with each week being a random subset of the total sample. The sub-sample was chosen by randomly selecting 19 weeks from the original field work schedule.

Each week contained seven Primary Sampling Units (PSUs) for a total of 133 PSUs. In each PSU the teams were to interview 12 of the original 15 households, with the remaining three to serve as replacements. The total nominal sample size was thus 1596.

Additional interviews

Following the collection and initial analysis of the data, it was determined that data from one district, Manatuto, and partially from another district, Oecussi, were of insufficient quality in certain modules. Therefore it was decided to repeat the survey in another 25 PSUs of these two districts – six in Manatuto, and 19 in Oecussi. The additional PSUs chosen were randomly selected within the two districts from the remaining non-panel PSUs in the original TLSLS2 sample.

	original	additional	total
Aileu	1	0	1
Ainaro	13	0	13
Baucau	12	0	12
Bobonaro	10	0	10
Covalima	6	0	6
Dili	14	0	14
Ermera	12	0	12
Lautem	10	0	10
Liquica	9	0	9
Manufahi	7	0	7
Manatuto	6	6	12
Oecussi	27	21	48
Viqueque	5	0	5
Total	132	27	159

Weights

Due to the necessity of additional interviews, there are three possible combinations of the data, with each combination having its own set of weights:

- (1) the original extension data
- (2) the original data, excluding the “questionable” data and including the additional interviews

[†] The TLSLS2 survey included both cross-sectional and panel components. As the structure of these two aspects already prohibited the pooling of the information for analysis purposes, the panel households were excluded from the extension sample selection. See documentation on TLSLS2 for notes on the selection of the original sample.

(3) the complete data, including both all the original data and the additional interviews

Therefore three different sets of sampling weights have been calculated.

The sample weights for the extension survey are indirect weights based on the original probability weights calculated for the TLSLS2. The TLSLS2 weights were calculated by Juan Muñoz of Sistemas Integrals. These weights were based on each household's selection probability, and then scaled by an adjustment factor, intended to match the demographic projections for the population of urban and rural areas in the five major regions of Timor Leste in mid-2007.[‡]

Region	Districts	Population	
1	Baucau, Lautem and Viqueque	urban	22,546
		rural	223,479
2	Ainaro, Manufahi and Manatuto	urban	24,020
		rural	128,474
3	Aileu, Dili and Ermera	urban	193,491
		rural	171,350
4	Bobonaro, Cova Lima and Liquica	urban	25,973
		rural	192,805
5	Oecussi	urban	9,244
		rural	56,251

As the extension survey was selected as a sub-sample of the original TLSLS sample, the original weights were used as a basis for the construction of the extension weights. In consultation with Juan Muñoz, it was decided to also use adjustment factors, to have the population estimates of the extension survey match the demographic projections in urban and rural areas of the five regions, for each of the three possible combinations of data described above.

The indirect weights are constructed by determining a scaling factor for the original weights which would bring the estimated population of the ten strata up to the same level as the original population projections. Separate scaling factors were calculated for each of the three possible combinations of the data. (Scaling factors are included in weights provided in the dataset and do not need to be added to analysis.)

		f1	f2	f3
Region 1	urban	2.390	2.390	2.390
	rural	3.355	3.355	3.355
Region 2	urban	2.137	2.157	2.020
	rural	3.244	3.085	2.286
Region 3	urban	3.514	3.514	3.514
	rural	3.104	3.104	3.104
Region 4	urban	2.221	2.221	2.221
	rural	3.298	3.298	3.298
Region 5	urban	1.723	2.553	1.309
	rural	3.561	1.903	1.306

[‡] The population projections were provided by Ricardo Neupert to Juan Muñoz in a private communication.

For the above, the scaling factors are constant across the three sets of weights for regions 1, 3 and 4 – those unaffected by re-interviews. Region 2 shows a small change between factors 1 and 2, based on the slightly different composition of the households but having the same sample size. There is a decrease between factors 1 and 2 and factor 3 as the third combination of the data includes both the original and re-sampled households, therefore effectively over-sampling region 2. The reduction in the scaling factor and by extension the weights corrects estimates for this over-sampling.

Region 5 (Oecussi) shows a large change between factors 1 and 2, in addition to the expected decrease due to over-sampling in factor 3. This large change results from the difficulty of exactly reproducing the original stratification of the PSUs into urban and rural areas in Oecussi. Therefore urban households in Oecussi were effectively over-sampled during the original extension, then under-sampled during the re-interviews. Because of the resulting large differences in factors 1 and 2, it was decided that in Oecussi alone, it would be more logical to calculate the adjustment factors based on the population as a whole rather than including the urban/rural stratifications.

Therefore the scaling factors used to calculate the extension weights are as follows:

Scaling Factors

		f1	f2	f3
Region 1	urban	3.355	3.355	3.355
	rural	2.390	2.390	2.390
Region 2	urban	3.244	3.085	2.286
	rural	2.137	2.157	2.020
Region 3	urban	3.104	3.104	3.104
	rural	3.514	3.514	3.514
Region 4	urban	3.298	3.298	3.298
	rural	2.221	2.221	2.221
Region 5	all	3.095	1.974	1.307

A short check was also performed to compare one parameter in Oecussi, household size, under the three different weighting systems, with the original estimates from the complete TLSLS2 data.

Weighted Estimates of Household Size in Oecussi

	Variable	Obs	Mean	Std. Dev.
original	rural	419	4.70	2.07
	urban	299	5.54	2.31
w1	rural	156	4.51	2.06
	urban	156	5.91	2.47
w2	rural	180	4.86	2.25
	urban	132	5.89	2.63
w3	rural	312	4.75	2.19
	urban	228	5.80	2.41

None of the estimates from the extension data with the indirect weights are statistically different from the original estimates based on the TLSLS2 data, nor are these estimates statistically significantly different from each other.

Organization and Timing of Survey

Field work for the TLSLS2-X was carried out by field teams from the Direcção Nacional de Estatística [DNE], with training and supervision provided by the World Bank, and Mekong Economics and Sistemas Integrales, private consulting firms. Each field team consisted for three interviewers, a supervisor, a data entry operator, and a driver. Data entry was concurrent with data collection, and performed in the field using laptop computers.

The questionnaire was developed by the individual topic teams within the World Bank in the fall of 2007. Pilot testing was conducted in January 2008. Unforeseen political events delayed the start of the training until May 2008. Ten days of training was conducted between May 13 and May 23 in Dili. Training was conducted by Sistemas Integrales with assistance from the World Bank, and consisted of both classroom exercises and field training. Field exercises were conducted in Dili, Alieu and Liquica districts.

Field work was originally scheduled for 20 weeks beginning in May and being completed in August. Questions as to the quality of the data arose during the compiling and cleaning process in September 2008. A “spot-check” data quality review mission was conducted in October 2008, and at that time it was determined that further interviews would be necessary in two districts. The additional interviews were conducted in November and December 2008. Data cleaning and compilation took place in January and February 2009, with the finished dataset being released to World Bank team members in February 2009. Plans are still on-going for public dissemination of the data.

Distribution of Field Work (by Teams)

Team	1	2	3	4	5	6	7
TLSLS2-X	Oecussi	Bobonaro Covalima Ermera	Ermera Liquica	Dili Oecussi	Ainaro Lautem Manufahi	Aileu Baucau Dili Manufahi Manatuto Viqueque	Baucau Lautem

The teams were condensed and reorganized for the additional interviews in Oecussi and Manatuto.

Using the Data

Sampling Weights

As previously mentioned, there are three sets of sampling weights to be used depending on the use of “questionable” data and the re-interviews. “Questionable” and re-interview data are identified by indicator variables in each data file. (Note that re-interviews were done with new households in new PSUs, not in the originally selected households.) Weight variables are labeled as to which configuration to which they relate.

Data Files

Data are broken down into 21 individual dataset. The following table lists the files by section and unique identifier variables.

Section	Title	Data Sets	Identifying Variables
Section 0	Cover	00_xcover.dta	xhh_id
Section 1	Household Information	sect01_indi.dta	xhh_id xpid
		sect01_xQ27.dta	xhh_id
		sect01_xQ28_29.dta	xhh_id x01_bline
Section 2	Agriculture	sect02.dta	xhh_id
		sect02_forest.dta	xhh_id product
		sect02_plots.dta	xhh_id plot
Section 3	Finance	sect03a.dta	xhh_id
		sect03b.dta	xhh_id xpid
Section 4	Shocks and Vulnerability	sect04a.dta	xhh_id shock
		sect04b.dta	xhh_id shock
		sect04c.dta	xhh_id
		sect04d.dta	xhh_id program
Section 5	Justice for the Poor	sect05a1.dta	xhh_id
		sect05a2.dta	xhh_id conflict
		sect05b.dta	xhh_id
		sect05c1.dta	xhh_id ic
		sect05c2.dta	xhh_id dc
		sect05d1.dta	xhh_id disp_code
		sect05d2.dta	xhh_id x05d2_dc
		sect05d3.dta	xhh_id
---	Weights	weights.dta	xhh_id

Potential Data Quality Issues

Agriculture

Similarly to the individual roster of the previous section, the plots listed in the previous survey are listed on the pre-printed cover page and all changes noted. The agricultural section, similarly to the other sections, suffers from problems with open-ended questions. This is particularly the case for the question asking what community restrictions are placed on the clearing of forest land (section 2d). The translation from the original question was vague (using the Tetun word for “boundary” for “restriction,”) and therefore many of the responses relate to physical boundaries on the land, such as stone walls and tree lines. Additionally, the translation of all answers from Tetun into English is imperfect, and those wishing to use this information for

analytical purposes are advised to also refer to the original Tetun. Analysts should be careful in using the data from the open ended questions because of translation problems.

Also, it was noted during the training and field work that many interviewers had significant difficulties understanding definitions with some of the land management and investment questions. In general, however, all agricultural data may be used for analysis, sampling weights $w3$.

Finance

It should be noted that the quality of the data for the finance experiment (comparing the knowledge of the household head to that of other household members) was not sufficient for the experiment to be deemed a success. Subsequent spot-checking revealed that in many cases, interviewers asked the household head *about* the financial activities of various household members instead of asking them directly. Therefore this data should only be used to measure the access to finance at the household level. The finance sections were not repeated during the additional interviews in the replacement PSUs. Sampling weights $w1$ should be used when doing any analysis with this data.

Shocks and Vulnerability

It was determined following the initial round of data collection that the shocks and vulnerability module had some issues with uneven interview quality. Two reasons were listed as potential causes of the data quality issues: (1) fundamental inability to adequately translate both the word and concept of a “shock” into the Timorese context, and (2) incomplete / questionable responses to the health shock questions in particular. Analysis for health shocks should drop the “questionable” households and use the “re-interview” households, sampling weights $w2$.

Justice for the Poor

Similar to the shocks and vulnerability module, the justice module included a long series of follow up questions if the household indicated having experienced a dispute during the recall period. Again, the number of disputes experienced by the household seemed extremely low compared to expectations. This was particularly a problem with the Manatuto district in which no disputes were recorded during the first set of TLSLS2-X interviews. Analysis for the disputes section of the justice module should drop the “questionable” households and use the “re-interview” households, sampling weights $w2$.

The justice model also has a number of instances in which the specifications for “other” were not recorded. Every effort was made to ensure this data was as complete as possible, but gaps do remain.

Also, data users should use caution when using the **imputed rank** variable in section 5D. The rank in terms of importance was not explicitly captured in the data entry software, and the rankings therefore had to be imputed from the order they were listed in the original data entry. Inconsistencies may exist in this variable.

Data Cleaning

The TLSLS2-X had a significant number of responses in which the response is “other”. In general, if the response clear fit into a pre-coded response category, it was recoded into that

category during the cleaning and compilation process. Some responses where additional information was provided were not recoded even though they clearly fit into pre-coded categories. For example, “agriculture project” would be recoded into the “agriculture” category, while “community garden” would not. Data users can either use the additional information, or re-code into categories as they see fit.

Appendix A : Basic Sample Size Information

Completed Interviews by District[§]

District	Completed	Partially	Not completed	Total
Aileu	12	0	0	12
Ainaro	153	2	8	163
Baucau	143	1	3	147
Bobonaro	120	0	7	127
Covalima	72	0	3	75
Dili	154	2	20	176
Ermera	144	0	7	151
Lautem	120	0	3	123
Liquica	108	0	9	117
Manufahi	79	5	6	90
Manatuto	113	31	9	153
Oecussi	512	28	51	591
Viqueque	59	1	0	60
Total	1,789	70	126	1,985

Questionable and Re-Interviews by District (of Completed Interviews)

District	Completed	Questionable	Re-Interviews	Total
Aileu	12	0	0	12
Ainaro	153	0	0	153
Baucau	143	0	0	143
Bobonaro	120	0	0	120
Covalima	72	0	0	72
Dili	154	0	0	154
Ermera	144	0	0	144
Lautem	120	0	0	120
Liquica	108	0	0	108
Manufahi	79	0	0	79
Manatuto	0	58	55	113
Oecussi	84	205	223	512
Viqueque	59	0	0	59
Total	1,248	263	278	1,789

[§] "Completed" means that the interviewer asked all sections in accordance with the instructions provided in the interviewed manual, not that all sections were completed. For example, if there was no spouse present in an "odd" household, the justice section was not asked. The interview would still be marked as "completed."

Sample Size by Section

District	Section 1 (individuals)	Section 2 (households)	Section 2 (plots)	Section 3 (households)	Section 3 (individuals)	Section 4 (shocks)	Section 4 (households)	Section 5 (households)	Section 5 (conflicts)	Section 5 (disputes - incidence)	Section 5 (disputes - resolution)
Aileu	71	12	34	12	25	198	11	11	44	99	1
Ainaro	904	153	325	163	356	2,718	151	142	588	1,323	33
Baucau	782	143	210	147	299	2,430	138	137	552	1,242	21
Bobonaro	597	120	176	127	252	2,070	115	107	452	1,017	88
Covalima	381	72	79	75	141	1,260	70	65	260	585	29
Dili	1,066	154	178	176	444	2,682	149	142	584	1,323	44
Ermera	853	144	490	151	322	2,448	136	131	524	1,179	105
Lautem	726	120	168	123	274	2,124	118	115	460	1,035	8
Liquica	688	108	317	117	231	1,944	108	101	420	945	48
Manufahi	498	79	155	90	177	1,440	80	77	308	693	9
Manatuto	754	113	168	73	143	2,430	135	127	508	1,143	16
Oecussi	2,779	512	840	333	626	9,306	517	502	2,016	4,527	85
Viqueque	333	59	71	60	125	1,044	58	57	228	513	14
Total	10432	1,789	3,211	1,647	3,415	32,094	1,786	1,714	6,944	15,624	501

Appendix B : Available Documentation

Basic Information Document TLSLS2-X
Survey Manuel TLSLS2-X
Households Survey TLSLS2-X (English & Tetun)

Basic Information Document TLSLS2
Survey Manuel TLSLS2
Households Survey TLSLS2 (English & Tetun)