



**KINGDOM OF CAMBODIA
NATION RELIGION KING**

Basic Information Document

**Cambodia
Living Standards Measurement Study - Plus (Cambodia LSMS+)**

**National Institute of Statistics
Ministry of Planning
Phnom Penh, Cambodia**

ACRONYMS

CAPI	Computer Assisted Personal Interview
LSMS	Living Standards Measurement Study
LSMS-ISA	Living Standards Measurement Study – Integrated Surveys on Agriculture
LSMS+	Living Standards Measurement Study - Plus
NIS	Cambodia National Institute of Statistics
WB	World Bank

Table of Contents

Contents

Introduction.....	4
The Survey Instruments	4
Baseline Sample and Weights.....	5
3.1 Survey Sample	5
3.2 Survey Weights	7
Training of Interviewers for the Survey.....	9
Field Work.....	9
5.1 Organization of Fieldwork	9
5.2 Pre-loaded Information	Error! Bookmark not defined.
5.3 Respondents	Error! Bookmark not defined.
Data Management and Description of Datasets.....	11
6.1 Data Management.....	11
6.1.1 CAPI	Error! Bookmark not defined.
6.1.2 Data Communication System.....	11
6.1.3 Data Cleaning.....	11
6.2 Description of Datasets.....	11
Table 6-1: Datasets.....	12
6.2.1 Confidential information	13
Using the Data	14
7.1 File Structure.....	14
7.2 Merging Datasets	14

1.0 Introduction

The purpose of this document is to provide detailed information on the Cambodia Living Standards Measurement – Plus Survey 2019-20.

As part of the IDA18 Gender and Development Window, the World Bank has made a corporate commitment to collect self-reported, individual-level, intra-household information on asset ownership, employment and entrepreneurship in 6 IDA countries.¹ To fulfill the corporate commitment to collect such data, the World Bank Living Standards Measurement Study (LSMS) Plus program has been established – anchored in the partnership between the Development Data Group and the Gender Group.

The objective of the LSMS Plus program is to integrate the World Bank-required individual-level data collection into upcoming household surveys in IDA countries that do NOT otherwise have an official mandate to collect such data. This mode of integration increases not only the cost effectiveness of Bank-financed data production (compared to implementing stand-alone surveys), but also the operational value of Bank-financed individual-level data that would be collected as part of multi-topic surveys that also collect detailed information on monetary and non-monetary aspects of welfare.

The Cambodia LSMS Plus Survey was conducted in 25 percent of the enumeration areas (EAs) that was visited by the 2019/20 round of the Cambodia Socioeconomic Survey (CSES) which is implemented by the National Institute of Statistics (NIS).

Specifically, in the CSES EAs that were visited during the period of October-December 2019, the NIS selected additional households that were NOT be part of the planned CSES sample and that constituted the LSMS Plus Survey sample, to whom the World Bank-required survey questionnaire modules was administered, based on the World Bank-required fieldwork protocols. The LSMS Plus Survey data was collected in parallel with the CSES data in these EAs.

2.0 The Survey Instruments

The questionnaire has two parts. The first part is the Household Questionnaire. The second part is the Individual Questionnaire. The Individual Questionnaire was administered to all adult individuals (18 years and above) in the household

The design of all questionnaire modules was informed by the design of the questionnaire in other countries but have been adapted to the Cambodian context.

In Cambodia, each adult household member was administered the LSMS+ Survey individual-level modules on ownership of and rights to (i) land, (ii) apartments and condos, (iii) livestock, (iv) consumer durables and valuables, (v) financial assets, and (vi) mobile phones and valuables. The asset modules will ask separately regarding (i) reported ownership, (ii) economic ownership, (iii) documented ownership, (iv) perceived tenure security, and rights to (v) sell, (vi) bequeath, (vii) use as collateral, (viii) rent out, and (ix) make improvements/invest.

¹ For more information, please see: <http://documents.worldbank.org/curated/en/846661468196734005/pdf/106181-BR-IDA-18-Gender-Development-PUBLIC-disclosed-6-2-2016-11-31PM-IDA-SecM2016-0114-Box360260B.pdf>.

3.0 Survey Sample and Weights

3.1 Survey Sample

The survey took approximately 3 months running from October-December 2019 with a few final interviews in early January 2020. The NIS sampled 6 additional households in the 252 enumeration areas that were visited during the period of October-December 2019 by the 2019/20 round of the Cambodia Socioeconomic Survey (CSES). The LSMS+ Survey sample includes a total of 1,512 households.

3.2 Survey Weights

SAMPLING METHODOLOGY

1. The primary Sampling Units (PSUs)

The Primary Sampling Units (PSUs) of this survey were the subsamples of the selected PSUs of the Cambodia Socio-Economic Survey (CSES) 2019/20². The PSU in this case can either be a village (if the village is small) or an Enumeration Area (EA) from the mapping operation of 2019 General Population Census of Cambodia (if the village is large, probably exceed 120 households). These covered all the CSES' s sample villages in three months (those selected for interviews during the October - December period of fieldwork) out of its twelve-month sample. The number of sample villages and households in each province are shown in the following table:

Table 1: Number of sample villages and households in each province

	Province	No. of Sample Villages (PSU)	No. of Sample Households (SSU)
1	Banteay Meanchey	12	72
2	Battambang	18	108
3	Kampong Cham	12	72
4	Kampong Chhnang	12	72
5	Kampong Speu	12	72
6	Kampong Thom	10	60
7	Kampot	10	60
8	Kandal	16	96
9	Koh Kong	6	36
10	Kratie	6	36
11	Mondul Kiri	6	36
12	Phnom Penh	24	144
13	Preah Vihear	6	36
14	Prey Veng	12	72
15	Pursat	12	72
16	Ratanak Kiri	5	30
17	Siemreap	13	78
18	Preah Sihanouk	6	36
19	Stung Treng	6	36

² For more details on the CSES sample village selection, please see the CSES 2019-2020 report, NIS.

20	Svay Rieng	10	60
21	Takeo	14	84
22	Oddar Meanchey	6	36
24	Pailin	6	36
25	Tboung Khmum	12	72
TOTAL		252	1512

2. The Secondary Sampling Units (SSUs)

The SSUs in this survey constitute sample households. In this stage, 6 households were selected in each selected PSU. The selections of these households were carried out in the field by the field enumerators. The selection was done under the Circular Systematic Random Sampling (CSRS) scheme using the PSU frame of household from the household listing conducted by the CSES field enumerator in the selected PSU. The process of household selection is summarized as follows:

a). Let M^* be the number of households in the selected PSU, and the number of households to be selected in each village (n) is 6 then the sampling interval (I) is calculated as M^*/n and rounds it up to the nearest two decimal point.

b). Take a random start between 1 and M^* from the random number table together with the instruction provided in data collection manuals. Suppose that a random start was taken, say r then the set of selected households are:

$$r, r + I, r + 2I, r + 3I, \dots, r + (n-1)I.$$

Note: 1. If any $(r + (n-1)I) > M$, then take $[(r + (n-1)I) - M^*]$

2. If $(r + (n-1)I)$ has a decimal point, rounds it up the nearest integer.

3. The sampling interval (I) constitutes an equal jump from one sample to another throughout the PSU.

3. The Sampling Weight

3.1 Designed Weight

Stage 1: Selection of villages: The probability of village j in province i to be selected (P_{ij}) is represented by:

$$P_{ij} = \frac{n_i M_{ij}}{M_i}, \text{ where:}$$

- n_i denotes the number of villages in province i to be selected as the sample.
- M_i denotes the total number of households in province i according to the frame. That is; $M_i = \sum_{j=1}^{N_i} M_{ij}$, where N_i denotes the total number of villages in province i .
- M_{ij} denotes the number of households in village j of province i according to the frame.

The sampling weight for the stage 1 is represented by:

$$W_{ij}^1 = \frac{1}{P_{ij}} = \frac{M_i}{n_i M_{ij}}$$

Since the samples in this stage are the three-month subsamples of the twelve-month CSES 2019-2020 samples, this weight was derived as the first-stage weight from the CSES 2019-2020. The adjusted designed weight of this survey in this stage is represented by: $W_{ij}^{1adj} = 4 W_{ij}^1$.

Stage 2: Selection of 6 households in each village/PSU selected in the first stage: The probability of the household k to be selected in village j of province i (P_{ijk}) is represented by: $P_{ijk} = \frac{6}{M_{ij}^*}$, where:

- M_{ij}^* denotes the total number of household in village j of province i that was actually reported by the village chief in the data collection day. The weight for the second stage is represented by:

$$W_{ijk}^2 = \frac{1}{P_{ijk}} = \frac{M_{ij}^*}{6}$$

The designed weight (W_{ijk}^D) of the survey would be calculated as:

$$W_{ijk}^D = W_{ij}^{1adj} \cdot W_{ijk}^2 = 4 \left(\frac{M_i}{n_i M_{ij}} \right) \left(\frac{M_{ij}^*}{6} \right)$$

3.2 Sampling Weight Adjustment

The adjusted sampling weight for non-response (W_{ijk}^{adj}) was computed as:

$$W_{ijk}^{adj} = (W_{ijk}^D) \frac{6}{m_{ij}}, \text{ where } m_{ij} \text{ denotes the number of sample households to be actually interviewed}$$

in village j of province i . If there is no non-response, then m_{ij} would be equal to 6.

The above adjusted weight would then be calibrated to account for the differences between the total number of households reported in frame and the recently updated data on the total number of households which was available from the provisional results of the 2019 General Population Census of Cambodia in each province to be the final household weight. The calibration factor for each sample household is calculated as:

$$C_{ijk} = \frac{\text{No. of households from the 2019 General Population Census of Cambodia}}{\text{No. of households weighted by } W_{ijk}^{adj}}$$

The final household weight (W_{ijk}^{Final}) would then be derived as: $W_{ijk}^{Final} = W_{ijk}^{adj} * C_{ijk}$

The final person weight for male population is represented by:

$W_{ijkm}^{Final} = W_{ijk}^{Final} * C_{ijkm}$, whereas, the final person weight for female population is represented by:

$W_{ijkf}^{Final} = W_{ijk}^{Final} * C_{ijkf}$, where:

$$C_{ijkm} = \frac{\text{No. of male population from the 2019 General Population Census of Cambodia}}{\text{No. of male population weighted by } W_{ijk}^{Final}}, \text{ and}$$

$$C_{ijkf} = \frac{\text{No. of female population from the 2019 General Population Census of Cambodia}}{\text{No. of female population weighted by } W_{ijk}^{Final}}$$

4.0 Training of Interviewers for the Survey

The NIS, with technical support from the World Bank, trained the 42 LSMS+ Survey enumerators to carry out the fieldwork. The training included formal classroom training and field practice lasted 14 days. Given the complexity of the data collection set-up and the sensitive nature of the material contained in the individual-level data collection, emphasis was placed on content training, field practice, the use of the Survey Solutions computer assisted data entry application, and fieldwork preparations. Additionally, a training session was held on (i) respondent sensitization, (ii) building rapport, (iii) avoiding conflicts at the household and community levels in the context of soliciting sensitive information, (iv) tips for scheduling one-on-one interviews and (v) soliciting information on hidden assets was implemented.

5.0 Field Work

5.1 Organization of Fieldwork

One enumerator was assigned to each LSMS+ Survey household for the household-level questionnaire. Individual-level questionnaires were to be shared between enumerators so that male enumerators interview male respondents and female enumerators interview female respondents. During the fieldwork, the enumerators were expected to sync their tablets daily, and the NIS regularly monitored the completion and

quality of data collection, and made the enumeration area and household assignments, as part of the Survey Solutions system.

In each household, the LSMS+ Survey surveyed all adult individuals (18 years and above) in private, on (i) ownership of and rights to selected physical and financial assets, and (ii) health, education, and employment (iii) migration and (iv) time use. In addition, at least one of the managers of each household off-farm enterprise was interviewed as part of a module on household enterprises.

6.0 Data Management and Description of Datasets

6.1 Data Management

6.1.1 Data Entry Platform

To ensure data quality and timely availability of data, the Cambodia LSMS+ was implemented using the World Bank's *Survey Solutions* CAPI software.³ To carry out the survey, 1 laptop computer and a wireless internet router were assigned to each team, and each enumerator had a tablet computer. The use of *Survey Solutions* allowed for the real-time availability of data as the completed data was completed, approved by the Supervisor and synced to the Headquarters server as frequently as possible. While administering the first module of the questionnaire the enumerator(s) also used their tablets to record the GPS coordinates of the dwelling units.

6.1.2 Data Management

The *Survey Solutions* CAPI based data entry application was designed to stream-line the data collection process from the field. Interviews were collected in “sample” mode (assignments generated from headquarters) as opposed to “census” mode (new interviews created by interviewers from a template) for the NSS to have more control over the sample.

The range and consistency checks built into the application were informed by the LSMS+ experience in other countries and by other surveys conducted in Cambodia. Prior programming of the data entry application allowed for a wide variety of range and consistency checks to be conducted and reported and potential issues investigated and corrected before closing the assigned enumeration area. Headquarters (the NSS management) assigned work to the supervisors based on their regions of coverage. The supervisors then made assignments to the enumerators linked to their supervisor account. The work assignments and syncing of completed interviews took place through a Wi-Fi connection to the server. Because the data was available in real time it was monitored closely throughout the entire data collection period and upon receipt of the data at headquarters, data was exported to Stata for other consistency checks, data cleaning, and analysis.

6.1.3 Data Cleaning

The data cleaning process was done in two main stages. The first stage was to ensure proper quality control during the fieldwork. This was achieved in part by incorporating validation and consistency checks into the *Survey Solutions* application used for the data collection and designed to highlight many of the errors that occurred during the fieldwork.

The second stage of cleaning involved a comprehensive review of the final raw data following the first stage of cleaning. Every variable was examined individually for (1) consistency with other sections and variables, (2) out of range responses, and (3) formatting.

6.2 Description of Datasets

³ For background and documentation on *Survey Solutions*, please visit <https://mysurvey.solutions/>. The software platform is available free of charge and is being developed by the World Bank Development Data Group - Data Analytics and Tools Unit (DECAT). To access *Survey Solutions* Designer, please visit and sign up as a user at <https://designer.mysurvey.solutions/>. All IHS5 CAPI questionnaires are available free of charge to any interested implementing agency.

These modules were administered in the Cambodia LSMS+

- Household Roster
- Children Living Elsewhere
- Housing
- Food, Beverage, and Tobacco Consumption during the last 7 days
- Recall Non-food Expenditures
- Household Enterprises
- Land Ownership at the Household-Level
- Apartment/Condo Ownership at the Household-Level⁴
- Livestock Ownership at the Household-Level
- Consumer Durables Ownership at the Household-Level
- Education and Literacy
- Health
- Migration
- Employment
- Time Use
- Land Ownership at the Individual-Level
- Apartment/Condo Ownership at the Individual -Level
- Livestock Ownership at the Individual -Level
- Consumer Durables Ownership at the Individual -Level
- Mobile Phones
- Financial Assets

Table 6-1 show the sections of the Questionnaire and their corresponding datasets.

Table 6-1: Datasets

Section	Section Name	Dataset Filename
1	Interview Information	HH_SEC_1.dta
2	Household Roster	HH_SEC_2.dta
3	Children Living Elsewhere	HH_SEC_3.dta
4	Housing	HH_SEC_4.dta
5	Food Consumption	HH_SEC_5.dta
6	Non-food Consumption	HH_SEC_6.dta
7A	Household Enterprises Filter	HH_SEC_7A.dta
7B	Household Enterprises Characteristics	HH_SEC_7B.dta
7B	HH Enterprises Characteristics Household Labor	HH_SEC_7B_hhlabor.dta
7B	HH Enterprises Characteristics Non-HH Labor	HH_SEC_7B_nonhhlabor.dta
7C	Household Enterprises Expenditures	HH_SEC_7C.dta
7D	Household Enterprises Revenues	HH_SEC_7D.dta
8	Household-level Land Roster	HH_SEC_8.dta
9	Household-level Livestock Roster	HH_SEC_9.dta

⁴ Although there was a module on apartments and condo ownership that was administered to households, zero respondents reported owning this type of asset so the household- and individual-modules on this were removed from the questionnaire and data.

10	Household-level Durables	HH_SEC_10.dta
	Individual Cover	Individual_Cover.dta
11	Education and Literacy	HH_SEC_11.dta
12	Health	HH_SEC_12.dta
13A	Migration Internal	HH_SEC_13A.dta
13B	Migration International	HH_SEC_13B.dta
14	Employment	HH_SEC_14.dta
15	Time Use	HH_SEC_15.dta
16A	Land Ownership at the Individual-Level	HH_SEC_16A.dta
16B	Land Ownership Rights Holders	HH_SEC_16B.dta
16C	Land Ownership Other Rights Holders	HH_SEC_16C.dta
17A	Livestock Ownership at the Individual -Level	HH_SEC_17A.dta
17B	Livestock Ownership at the Individual -Level	HH_SEC_17B.dta
18A	Consumer Durables Ownership at the Individual -Level	HH_SEC_18A.dta
18B	Consumer Durables Ownership at the Individual -Level	HH_SEC_18B.dta
19A	Mobile Phones Roster	HH_SEC_19A.dta
19B	Mobile Phones Details	HH_SEC_19B.dta
20A	Financial Assets Roster	HH_SEC_20A.dta
20B	Financial Assets Details	HH_SEC_20B.dta

6.2.1 Confidential information

Note that, for purposes of maintaining the confidentiality of the data, all names, phone numbers, and addresses have been removed from the datasets.

7.0 Using the Data

7.1 File Structure

The data should always be used in conjunction with the questionnaire and the interviewer's instruction manual. Where there are no issues of confidentiality, all the variables from the questionnaire have been included in the data sets. In some cases, there is an additional variable which contains the "other specify" information that was written in the questionnaire. So, for example, the other specify variable was indicated with an "_oth" attached to the variable name such as s02q05_os containing the "other specify" information for a variable q5.

7.2 Merging Datasets

All datasets contain a variable (***HHID***), which is a unique identifier for the household. This variable is used as the unique key variable in the merging of all household type datasets. In some of the other types of datasets, additional key variables may be required in the merging process. In the case of individual type files, the variable that uniquely identifies the individual in the household is ***PID***. In order to merge any two individual type files, both the variables ***HHID*** and ***PID*** would be used.