



KINDOM OF CAMBODIA

Cambodia
Socio-Economic Survey
2004

Technical Report on Survey
Design and Implementation

National Institute of Statistics, Ministry of Planning

Phnom Penh, Cambodia
September 2005

The Cambodia Socio-Economic Survey 2004 has been carried out by the National Institute of Statistics, Cambodia, and with Technical Assistance provided by Statistics Sweden. The survey is funded by UNDP, SIDA and in specific areas sponsored by The World Bank

Table of Content

<u>FOREWORD</u>	2
<u>PREFACE</u>	3
<u>1 GENERAL</u>	4
<u>1.1 INTRODUCTION</u>	4
<u>1.2 OVERALL OBJECTIVES</u>	4
<u>1.3 SCOPE OF THE SURVEY</u>	5
<u>1.4 SURVEY ORGANISATION</u>	8
<u>2 SURVEY PLANNING</u>	8
<u>2.1 SAMPLING DESIGN</u>	8
<u>2.2 QUESTIONNAIRE DESIGN</u>	11
<u>3 FIELD OPERATIONS AND TRAINING</u>	13
<u>3.1 PILOT SURVEY</u>	13
<u>3.2 ENUMERATOR AND SUPERVISOR TRAINING</u>	13
<u>3.3 FIELD OPERATIONS</u>	14
<u>3.4 MONITORING</u>	16
<u>4 DATA PROCESSING</u>	16
<u>4.1 FIRST ASSESSMENT</u>	16
<u>4.2 DATA MANAGEMENT</u>	16
<u>4.3 TRAINING</u>	17
<u>4.4 DATA EDITING AND CODING</u>	18
<u>4.5 DATA ENTRY</u>	18
<u>4.6 REFERENCE DATABASE</u>	18
<u>5 TABULATION AND DISSEMINATION</u>	19
<u>REFERENCES</u>	20
<u>ANNEX 1. NIS SURVEY STAFF AND EXPERT ADVISORS</u>	21
<u>ANNEX 2. SAMPLING DESIGN AND ESTIMATION</u>	23
<u>ANNEX 3. SAMPLE ALLOCATION TO STRATA</u>	28
<u>ANNEX 4. TEAM ALLOCATION</u>	29
<u>ANNEX 5. FIELD SUPERVISION PLAN</u>	32
<u>ANNEX 6. EXCERPT FROM THE DATA MANAGEMENT MANUAL</u>	33
<u>ANNEX 7. EXCERPT FROM THE DATA EDITING AND CODING MANUAL</u>	37
<u>ANNEX 8. EXCERPT FROM THE DATA ENTRY MANUAL</u>	40
<u>ANNEX 9. ADJUSTMENT OF THE SAMPLING WEIGHTS</u>	45

HOUSEHOLD LISTING FORM

VILLAGE QUESTIONNAIRE
HOUSEHOLD QUESTIONNAIRE
DAIRY SHEETS
TIME USE FORM

FOREWORD

I have the pleasure to provide the foreword to the *Technical Report on Survey Design and Implementation of the Cambodia Socio-Economic Survey (CSES) 2004*. The CSES 2004 is the fifth Cambodia Socio Economic Survey conducted by National Institute of Statistics, since the Socio Economic Surveys in the years 1993/94, 1996, 1997 and 1999.

The CSES 2004 survey was designed to be the most comprehensive survey of its kind in the country. The main objective of the survey is to collect accurate statistical information about living standards of the population and the extent of poverty as an essential instrument to assist the government in diagnosing the problems and designing effective policies for reducing poverty, and in evaluating the progress of poverty reduction which are the main priorities in the "Rectangular Strategy" of the Royal Government of Cambodia.

The survey was planned and designed by the staff of NIS with overall technical and management assistance provided by Statistics Sweden, and with UNDP and SIDA funds. The World Bank has also been contributing in various ways. The CSES 2004 has filled data gaps in several subject matter areas, which will continue to meet the data needs of many users.

The report presents information on survey planning and implementation, organization and fieldwork, monitoring, survey methodology, data processing, and dissemination. All questionnaires are attached as a special annex.

The survey results on eight major subject matter areas are described in a separate report titled "*Summary Subject Matter Report*".

On behalf of the Royal Government of Cambodia, I would like to highly appreciate the significant financial and technical assistance provided by various international and bilateral donors, in particular, the UNDP, The World Bank, Sida, and Statistics Sweden, making the survey a success and meeting the highest international standards.

Chhay Than
Senior Minister
Minister of Planning

Ministry of Planning
Phnom Penh, Cambodia
September 2005

PREFACE

The National Institute of Statistics (NIS) of the Ministry of Planning has conducted the fifth Cambodia Socio-Economic Survey in the period 2003-2004 (CSES 2004). Previous surveys were undertaken in 1993/94, 1996, 1997 and 1999. The United Nations Development Program (UNDP), The World Bank and the Swedish International Development Agency (Sida) have been the main sponsors while Statistics Sweden has provided the Technical Assistance.

The main objective of the survey was to collect statistical information about living standards of the population and the extent of poverty. Essential areas as household production and cash income, household level and structure of consumption including poverty and nutrition, education and access to schooling, health and access to medical care, transport and communication, housing and amenities and family and social relations. For recording expenditure, consumption and income the *Diary Method* was applied for the first time. The survey also included a Time Use Form detailing activities of household members during a 24-hour period.

The CSES 2004 was carried out on a nationwide representative sample of 15,000 households within 900 Primary Sampling Units (PSUs). It was divided into 15 monthly samples of 1,000 households each in 60 PSUs from November-03 till January-05.

NIS decided to use statistical methods (calibration) to achieve better comparability between the different rounds of the CSES surveys by adjusting the samples to the population size and structure that was established by the national population census carried out in 1998. To mirror the rapid changes in the population, it proved necessary to project the population forwards to 2004 and backwards to 1993/94, taking into account fertility, mortality and internal migration rates.

The CSES 2004 is a very comprehensive survey that will provide a lot of statistical data to be used for various purposes. The primary data files will be made available to analysts at other ministries for further analysis, and to international organizations and university researchers according to the procedures specified in the 2005 Statistics Law.

We thank UNDP, WB and Sida for sponsoring the survey, and Statistics Sweden for management assistance and technical support. Furthermore, we wish to place on record our deep appreciation of the work carried out by project experts and consultants, NIS staff, staff of provincial planning departments and the staff of the Ministry of Planning who worked with dedication and enthusiasm to make the survey a success. We also extend our thanks to all the participating households and individuals.

San Sy Than
Director General, NIS

Phnom Penh, Cambodia
September, 2005

1 General

1.1 Introduction

A number of national household surveys have been conducted by the National Institute of Statistics (NIS) in the last ten years: four rounds of the Cambodia Household Socio-Economic Surveys (CSES) in 1993/94, 1996, 1997 and 1999, the 1996 Demographic Survey, the Population Census of Cambodia in 1998, the Labor Force Survey in 2000 and 2001, the Cambodia Demographic and Health Survey (CDHS 2000), the 2001 Child Labor Survey, and the 2004 Intercensal Population Survey.

The UNDP and The World Bank as well as other international organizations are actively supporting the poverty reduction strategy of the Royal Government of Cambodia, particularly by assistance to poverty monitoring and assessment. Still it has not been possible to establish a comparable time series on poverty with conclusions on poverty trends that have been trusted by all parties, including the Government. The conclusions of the comparison between the results of the 1997 and the 1999 surveys have remained especially controversial.

In 2002 it was decided to carry out a fifth round of the CSES in 2003/04 funded by UNDP. Statistics Sweden was contracted to provide the technical assistance.

NIS decided to establish new baseline data on household consumption and production both for poverty assessment and for the National Accounts by adopting the *diary method* for the 2003/04 round. To make it possible to analyze poverty trends it was also decided to include recall modules, thus in parallel use the *recall method* on food and non-food consumption.

It was initially planned to start the fieldwork in October 2003. Due to a postponed pilot survey and unexpected delays in the questionnaire design and field tests, the first survey month became November 2003, reducing the survey to 14 months. In late 2004 it was decided to add January 2005 retaining the original scheme of 15 survey months.

During the time span of the survey a large number of survey reports have been issued covering all areas of concern. They are listed in the references. Intermediate results have been published and discussed in NIS arranged workshops. In the latest workshop in June 2005 eight important subject matter reports were presented. These reports are the cornerstones of the *Summary Subject Matter Report* that has been issued recently.

The present *Technical Report on Survey Design and Implementation* should be regarded as a complement to the *Summary Subject Matter Report*, and as a record of the survey methodology and activities to serve planning and undertaking similar surveys in the future.

Throughout this report the survey, planned in 2003, carried out in 2004, and analyzed in 2005 will be referred to as the CSES 2004.

1.2 Overall objectives

Poverty reduction is a major commitment by the Government of the Kingdom of Cam-

bodia. Accurate statistical information about the living standards of the population and the extent of poverty is an essential instrument to assist the Government in diagnosing the problem, in designing effective policies for reducing poverty and in monitoring and evaluating the progress of poverty reduction.

To this end it is useful to consider the CSES 2004 as a means to strengthen the three interconnected subsystems of the emerging national statistical system of Cambodia; the population and social statistics system (counting people as to their living conditions), the economic statistics system (counting money as representations of the value created and used by various human activities) and the natural resources statistics system (monitoring the transformation and exploitation of natural resources).

Since economic progress most of the time is a precondition for social progress, economic statistics frequently have the center stage in public attention. Main economic indicators such as GDP growth, inflation rate, unemployment rate, balance of payments often lead to governmental action that will affect economic conditions of all households and businesses in the economy.

The National Accounts and their underlying economic statistics are designed to give a detailed but also a summary picture of a nation's economy as a basis for assessing the nation's economic situation and whether trends indicate improvement or deterioration in the national economy. In simple terms, the System of National Accounts is the book-keeping of a nation, recording the production and the income of economic actors in the different sectors of the economy, how this income is used for public and private consumption and investment as well as the value of the nation's productive assets.

The CSES should support the system of economic statistics that is the basis for modern National Accounts by providing much needed data on value added in household (informal sector) production. CSES data will make it possible to estimate the GDP share of private consumption from the use side or alternatively in terms of household production's share of the GDP from the production side.

A first principle for statistical system building is to look upon any survey as a follow-up of previous surveys to measure change over time as well as a preparation for future rounds of the survey. The information value of any survey *for the present* will increase the more it builds on previous rounds of the survey by enabling reliable comparisons over time to inform of progress made. The information value of any survey *for the future* will increase the more it sets a stable baseline for future surveys against which progress can be measured in a more reliable and relevant way. So obviously, there is a trade-off between a replication design of a new survey and a changed design for the benefit of the future.

A replication design should always be the preferred option for the living conditions modules in the previous surveys and a changed design an option to be considered for the measurement of the level and structure of household consumption and of household (informal sector) production. This is why it has been decided to substantially reduce the use of the recall method in favour of the diary method for capturing household consumption, production and income.

1.3 Scope of the survey

It was recognized that comparability between the CSES 2004 and the previous surveys must be ensured to the largest possible extent. To this end a list of modules and variables in the previous surveys was compiled and then used in the design phase. Briefly the four earlier CSES rounds have all made it possible to report sets of indicators on seven main areas of social concern:

1. Household production and cash income-earning activities by the labour force
2. Household level and structure of consumption, including poverty and nutrition
3. Education and access to schooling
4. Health (mother and child health in particular) and access to medical care
5. Transport and communication
6. Housing and amenities
7. Family and social relations, including gender issues and vulnerable groups

These seven areas were also covered by corresponding modules in the CSES 2004, the first two with the diary method as well as the recall method, the other five following the module design and variable content of previous rounds of the CSES with needed modifications and complements.

Population change at national, regional and local levels comes about through births, deaths, migration and family formation and dissolution. The 1998 population census of Cambodia is a major achievement re-establishing statistics of the population stock. The next stocktaking will be the population census planned for 2008. In the intercensal period there is a need to follow demographic developments closely. All the previous rounds of the CSES have registered the demographic characteristics of the population as to age, sex, marital status and migration. The 1997 and the 1999 rounds were the most complete as to demographic variables, including also births and deaths in the household in the last 24 months. The household sample size of the CSES 2004 will be too small to meet the needs for reliable estimates of births and deaths but will be large enough to capture the data on labor force by occupation and industry, and the school enrolment rates and educational levels of the population.

One way to get an overview of social statistics is to identify the broad sectors of social concern. Countries have implicitly agreed on what some of those concerns are in setting up universal agencies within the United Nations system. There is ILO for work, WHO for health, UNESCO for education and culture, FAO for food, Habitat for housing, IMF and World Bank for economic matters, Interpol for security against crime, UN itself for peace, etc. Most countries have ministries and policies that reflect these same concerns in their government structure.

Several of these international sector organizations have adopted lists of social concerns and recommended indicators to measure these concerns and which overlap to a large extent. To name a few, there are the Minimal National Social Data Set (the MNSDS), the Millenium Development Goals (MDG), and the Comprehensive Development Framework (CDF). There are also international activities based on Poverty Reduction Strategy Papers (PRSP).

The Royal Government of Cambodia approves of these efforts in e.g. adapting the MDG

and PRSP, and incorporating them into the “*Rectangular Strategy of Cambodia*”.

Cambodia is still a predominantly rural and agricultural society. The vast majority of the populations get their subsistence in households as self-employed in agriculture. The level of living is determined by the household’s command over labor and resources for own-production in terms of land and livestock for agricultural activities, equipments and tools for fishing, forestry and construction activities and income-earning activities in the informal and formal sector.

Monthly data to calculate value added in household production have been obtained from the household diaries as well as labor force input for compilation of GDP from the production side in the National Accounts.

To understand poverty in Cambodia is to understand the reasons for low productivity in household production. All the four CSES rounds have included some (small or large) module to capture household ownership of the material means of production. Variables from the previous surveys have been selected and complemented to get a more systematic coverage. The guiding principle has been to include the items that make a difference for the productivity and production potential of the household.

A Time Use module was included in the CSES 2004, its purpose to provide data on household member activities and also to be used for a better understanding of the situation of women and men in Cambodia.

Level and structure of household consumption determine the need satisfaction of the population. The low and fragile consumption level of most Cambodian households makes poverty, nutrition and food security into urgent social concerns. The household diaries capture household consumption including both bought and own-produced items for the poverty line calculations. The diaries give level and structure of household consumption to reweigh the basket for the Consumer Price Index and for the compilation of GDP from the use side in the National Accounts.

Cambodia has a very young population. More than 40 percent of the population is below 15 years of age. Close to 25 percent is of primary school age 6-14. Access to schooling and quality of schooling available is of great concern for the future of the country. The modules on school enrolment and level of education used in the two most recent rounds of the CSES needed little change since they are in line with international recommendations.

Cambodia still has high mortality and high morbidity in infectious diseases. Vaccination and other preventive medical programs as well as access to medical care are important social concerns. The health module used in the 1999 round of survey with a recall of illnesses for the last four weeks was considered adequate for capturing data for all household members age 5 and above. Illness is one of the potentially important causes of differences in productivity and poverty between households.

HIV/AIDS is a looming threat to the health of the Cambodian people. A HIV/AIDS module was therefore included.

Cambodia’s infrastructure is still weak. Improved infrastructure in transport and communication is needed to increase access to markets for agricultural and other products. This kind of data has mainly been captured in the village questionnaire.

Housing conditions, including access to drinking water and sanitation, has very high priority also as health and environmental concerns. The standard module used in the 1999 questionnaire was considered adequate.

Large households and female-headed households are characteristics of the Cambodian society. The economic level of members of households is a function of the dependency ratio and other characteristics of the household to which they belong. A module on household food security was adapted as relevant to the Cambodia situation. The module included questions on number of weeks of hunger or starvation in the preceding calendar year.

In Chapter 2.2 a list of the CSES 2004 modules is found while the questionnaires are attached as *ANNEX 10*.

1.4 Survey organisation

NIS formed a *Core Group* in January 2003 consisting of five NIS directors and headed by the DG. The Core Group was responsible for all survey planning and activities. The group was also engaged in establishing and carrying out monitoring schemes during the fieldwork, in arranging workshops for dissemination of CSES data, and for reporting to the Statistical Advisory Committee. It has been responsible for the allocation and utilization of funds and in solving logistical problems during the course of the survey.

As the most important part of the organization, 200 enumerators and 50 supervisors were recruited in early 2003 and were subject to intensive training leading up to the fieldwork. In 2004 an additional 28 enumerators were recruited and trained to replace those who by then had resigned.

A list of NIS survey staff and survey advisors is provided in *ANNEX 2*.

2 Survey planning

2.1 Sampling design

Overview

The sampling design and implementation was made in March 2003. A three-stage sample design was devised. Since NIS already had a master sample based on the Population Census 1998, consisting of 600 villages, it was used. But in order to reach the preferred number 900 the sample was extended to include additionally 300 villages.

In the *first stage* a sample of villages was selected in the head office. The villages were initially stratified into 45 strata (province*urban/rural). The villages were selected using systematic sampling with probabilities proportionate to size (PPS). The size measures used for the selection were number of households in the village according the 1998 Census. The resulting sample thus consisted of 900 villages, 600 villages in rural areas and 300 in urban areas.

In the *second stage* one Census Enumeration Area (EA or alternatively PSU) was selected randomly also in the head office. At the beginning of the fieldwork all

households in the selected EA were listed using a household listing form, and following internationally recommended procedures. A systematic sample of households was then drawn in a *third stage*. The third stage sample was 20 households in rural areas and 10 households in the urban areas.

Design work

The work on sample design was carried out in the following areas:

1. Estimation of sampling errors and design effects in the CSES 1999
2. Calculation of optimal sample size within primary sampling units
3. Sample size and sample allocation for CSES 2003

The work was done in a group of NIS staff in the form of expert assisted hands-on training in sampling design and calculation of sampling errors.

The sampling design, including estimation formulas, is presented in *ANNEX 3*. For the calculations STATA 6.0 was used. Calculations were made for the estimates on food and non-food consumption, possession of durables and the prevalence of illness.

The *Coefficients of Variation* (CV) on national level estimates are generally below 4 percent. The exception is the CV for total value of assets where there are rather high CVs especially in the urban areas, which should be expected.

The CVs are somewhat higher in the urban and rural domains but still generally below 7 percent. For the five zones the average CVs are in the range 5 to 13 percent with a few exceptions where the CVs are above 20 percent. For provinces the CVs for food consumption are 9 percent on average.

The sample take within Primary Sampling Units (PSU) was set to 10 households per PSU in the CSES 1999. Now when data on variances were available it was possible to make crude calculations of the optimal sample take within PSU. Calculations on some of the central estimates in the CSES 1999 show that the *design effects* in most cases are in the range 1 to 5.

Intra-cluster correlation coefficients have been calculated based on the design effects. These correlation coefficients are somewhat high. The reason is that the characteristics that are measured tend to be concentrated (clustered) within the PSUs. The optimal sample size within PSUs under different assumptions on *cost ratios* and intra-cluster correlation coefficients was then calculated. The cost ratio is the average cost for adding a village to the sample divided by the average cost of including an extra household in the sample.

In the CSES it was chosen to adopt a fairly low cost ratio due to the fact that the interview time per household is long. Under this assumption the optimal sample size is probably around 10 households per village for many of the CSES indicators.

Primary Sampling Units (PSU)

In previous surveys PSUs have been villages. It was decided to use village as PSU also for the CSES 2004 mainly because the communes were considered too large (and too few) to serve efficiently as PSUs. Another factor weighing in favor of villages was the fact that there already exists a master sample of villages at NIS.

The master sample consists of 600 villages (88 urban and 512 rural villages). The selection of villages was made with PPS sampling, hence facilitating an approximately self-weighting design with equal workloads in the villages. It was discussed whether a further stratification on 3-4 crude income-level strata should be done in urban Phnom Penh in order to secure a good spread of the sample over different income levels. It was decided not to do such stratification. Phnom Penh has a large sample (90 villages) selected with systematic sampling over a geographically ordered sample frame; this will in itself secure a reasonably good spread of PSUs.

The master sample is allocated over the strata proportionally to the total number of households in the strata. A problem with the master sample is that due to the proportional allocation the urban sample is too small to provide for good estimates in the urban domain. It was therefore decided to expand the sample to include 600 rural villages and 300 urban villages. The additions are shown in the table below.

Province	Urban	Rural	Province	Urban	Rural
01 Banteay	12	5	13 Preah Vihear	3	1
02 Battambang	17	6	14 Prey Veng	7	9
03 Kampong Cham	6	14	15 Pursat	7	3
04 Kampong	5	3	16 Rattanak Kiri	3	0
05 Kampong Speu	5	6	17 Siem Reab	14	5
06 Kampong Thum	8	5	18 Sihanouk Vile	18	0
07 Kampot	5	5	19 Stueng Treng	2	0
08 Kandal	6	9	20 Svay Rieng	3	4
09 Kaoh Kong	4	1	21 Takeo	5	7
10 Kracheh	10	1	22 Oddar Meanchey	3	0
11 Mondol Kiri	0	0	23 Krong Keb	3	0
12 Phnom Penh	63	4	24 Krong Pailin	3	0

The allocation over strata of the full sample of 900 PSUs is given in *ANNEX 4*.

Secondary Sampling Units (SSU)

The 600 villages in the master sample are divided in small segments containing approximately ten households each by using census enumeration area maps. As a consequence the boundaries of the segments would be difficult to identify in the field. There would be a risk that housing units constructed after the census will be missed when households are listed within segments during the fieldwork. It was therefore decided not to use the segments in the second stage sampling. The available options are in this situation either (a) to select households directly on stage in the village or (b) to use the enumeration areas as secondary sampling units. Selecting households directly would require a listing of all households in the village prior to the fieldwork. Such a listing would become time-consuming in large villages. It was therefore decided that enumeration areas would be used as SSUs, and that one enumeration area is selected within each sampled village.

Implementation

Villages were selected with a systematic PPS procedure within each stratum. For each sampled village one census enumeration area (EA) was selected. As the enumeration areas are roughly of the same size, the selection was done with equal probability sampling.

10 households were selected in each sampled village in the CSES 99. Calculations indicated that this sample size was close to optimum. Since the optimum is rather flat, the loss in efficiency from sample sizes of 12-15 is fairly small.

From a purely sampling efficiency point of view, a larger sample than 15 households per village should not be taken. However, factors relating to interviewers' security and well-being weighed in favor of having two interviewers per village in the rural areas. A workload of 10 households between the two interviewers in the village was considered too small. A workload of 15-20 households would be reasonable. All things taken together resulted in a sample of 10 households in urban areas (with one interviewer per village) and 20 households in rural areas.

The resulting sample consisted of 300 urban PSUs and 600 rural PSUs. From the urban PSUs 10 households were selected while 20 households were selected from rural PSUs. The sample thus contained 15000 households to be interviewed during 15 fieldwork months with 1000 different households each month.

2.2 Questionnaire design

The questionnaire is one of the first items in a strategy for quality control in data collection through surveys. Any piece of information to be collected must be formulated as a question so that all interviewers can be trained to read the questions in the same way. The questions must be formulated in such a way that all interviewers feel comfortable reading the questions aloud and that all respondents understand the questions in the same way. The layout of the questionnaire must be done so that the interviewer immediately understands how the respondent's answer should be recorded. A lot of work is normally needed to meet these requirements that are built into the process of communication in the interview situation. This is the kind of work in which final perfection is elusive and further improvements can always be made.

The initial work on questionnaire design resulted in a first draft prepared by NIS in early 2003. With expert assistance from Statistics Sweden in March the same year, a systematic walk-through question by question was done. A number of essential problems to be solved were then identified while errors or minor problems were attended to at once. At the end of the exercise some issues remained that were discussed at a meeting with users and stakeholders and then were referred to a larger group within NIS.

Another set of serious deficiencies in the questionnaires was recognized by a WB consultant in June 2003. At this stage the questionnaires included additional modules or questions demanded by stakeholders. It was suggested that the growing complexity of the whole survey would require more training of the fieldwork staff than was originally planned for. Major revisions of the questionnaires were also made at this point.

The pilot was carried out in June in two provinces with a sample of 870 households. When analyzing the survey, the NIS core team identified another number of problems

still to be tended to and agreed to some fairly substantial changes in the questionnaires that were used. The many interventions by various parties, already mentioned, with special interests had under way led to the inclusion of several modules and questions with different formats and standards for questions. A main thrust of the agreements by the NIS core team was to ensure comparability with previous rounds of the CSES and more uniformity of formats and standards in the questionnaires.

The proposals, which the NIS core team for the CSES 2004 did agree upon, were then implemented. Using the combined technical expertise at hand, improvements in formulations and in the formatting of all the questionnaires were made. In the beginning of August the household questionnaire was finished while diary forms, the time use form and the village questionnaire remained to be completed.

NIS was able to start work with the translation of the household questionnaire from English to Khmer in September. In the meantime, a World Bank consultant had started to review the existing supervisor and interviewer manuals.

According to the NIS time plan, translation and printing of the questionnaires and manuals had to be completed by early October for training of supervisors to start 6 October and of interviewers to start 13 October. The first group of fieldworkers could then be dispatched to the sampled villages all over Cambodia after completed training 25 October. The training cycle of the second group of supervisors and interviewers would be done in November for the fieldwork starting 1 December.

The resulting set contains five forms or questionnaires:

1. Household listing form
2. Village questionnaire
3. Household questionnaire
4. Dairy sheet
5. Time Use form

The *Household listing* was done prior to the sampling, and recorded household information on e.g. location, number of members, principal economic activity.

The *Village questionnaire* was used to gather basic common information on demography, economy, infrastructure, rainfall and natural disasters, education, health, retail prices, employment and wages, access to common property, sales prices of land, and recruitment of children for work.

The following modules were included in the *Household questionnaire*:

1. Basic household information
 - a. List of household members
 - b. Summary of presence in the household
 - c. Information on migration
 - d. Food consumption during the last 7 days (recall method)
2. Education and literacy
3. Housing
4. Household economic activities
 - e. Land ownership
 - f. Production of crops

- g. Cost of cultivation
- h. Hypothetical questions on rental and sales market
- i. Inputs and outputs of livestock raising activities
- j. Input and output from fish cultivation
- k. Inputs and outputs from forestry and hunting
- l. Inventory of household non-agricultural economic activities during the past 12 months
- 5. Household liabilities
- 6. Household income from other sources
- 7. Durable goods and other expenditures (partial non-food recall)
- 8. Construction activities in the past 12 months
- 9. Nutrition
- 10. Fertility and child care
 - m. Fertility history
 - n. Child feeding and vaccinations
- 11. Mortality
- 12. Health check of children
- 13. Current economic activity
 - o. Activity status during the past 7 days
 - p. Main occupation during the past 7 days
 - q. Employment and earnings during the last 12 months
- 14. Health
 - r. Illnesses during the past 4 weeks
 - s. Smoking
- 15. HIV/AIDS
- 16. Victimization

The *Dairy sheet* (diary method) recorded

- a. Expenditure and consumption of own-produced food
- b. Household income and receipts

The *Time Use form* was used to record main activities hour by hour during a 24 hour period. The activities were grouped into 22 categories e.g. Sleeping, Eating, School etc.

All questionnaires are found in *ANNEX 10*. Enumerator and supervisor manuals have not been attached. They can however be obtained at NIS.

3 Field operations and training

3.1 Pilot survey

It was initially scheduled that the first half of 2003 should include sampling design, questionnaire construction and training for a full scale pilot in May, data entry and analysis of the pilot survey in June and July, finalizing questionnaires and manuals, training in September and start of the main operation 1 October. However, the preparations for the pilot could not be finished on time so it had to be postponed for one month.

The pilot fieldwork in June was done in a sample of 20 villages in Phnom Penh and 30 villages in Kandal. In each of 37 rural EAs a sample of 20 households was used. In the 13 urban EAs the sample size was 10 households. Thus the pilot covered 870 households and about 5000 persons in total.

The main data collected were registered in the Household questionnaire and the Time Use form, which covered one month “diary” and one week “time use”. The coding and data entry was planned for July and the analysis for August. However, there was also a delay in the data entry that hampered the analyses. Since data entry was finished only in late September, the pilot could not be used as intended.

Cooperation of households in the rural areas was excellent, and the response rate was reported close to 100 percent even in Phnom Penh.

It was found that the household questionnaire used in the pilot was once again in need of some major revisions and the World Bank did propose two major complements, of which one was not tested in the pilot.

Another main conclusion was that the data edit and entry had to be accelerated and an improved data management system designed and implemented.

3.2 Enumerator and supervisor training

Initial training was provided during nine days for a group of 20-30 staff (not all were attending all the time). This training included a translation into Khmer of selected parts of the questionnaire, and a field test in a village outside Phnom Penh where the participants performed test interviews in 16 households. The experiences from this exercise were followed up during the course. The course also included general aspects on survey methodology and ways of controlling for errors. Many of the findings from this training served as input to later stages.

Prior to the start of the fieldwork intensive interviewer and supervisor training was carried out.

The 200 interviewers and 50 supervisors recruited were split into two groups, each consisting of 100 interviewers and 25 supervisors. The two groups later alternated so that the first group did their fieldwork during *odd* survey months (i.e. November, January, March ...) while the second group covered the *even* survey months (i.e. December 2003, February, April ...).

The training was designed with this in mind. The first group was trained in October 2003 while the second group was trained in November 2003 using premises at the NIS head office. Training of the first group was provided in English by a WB consultant and simultaneously interpreted in Khmer by the appointed NIS officer. The second group was trained by NIS only.

Common was that the supervisors were first trained during one week, and then jointly with their interviewers for two weeks.

Before all fieldwork months the group in turn was gathered at the NIS to walk through the questionnaire and manuals in order to correct errors that were detected during the briefing sessions or the monitoring operations, and to learn how to handle any changes

that were introduced to the survey instruments.

Training manuals are large and have not been attached. They can however be obtained at NIS.

3.3 Field operations

Interviewers and supervisors were initially divided into teams of five persons (one supervisor and four interviewers), making in total 50 teams for the fieldwork. Each month 25 teams were working in the field with a workload of 10 households per interviewer. In urban areas four PSUs were allocated to one team while in rural areas two PSUs were allocated. The fieldwork plan was designed in order to gather around 60 households monthly per team.

For a given month the team arrived in the village three days before the first day of the month to tend to preparatory tasks like discussing with village authorities, filling out the Household Listing Form, and thereafter sample those households to be interviewed.

The Village Form was filled out by the supervisor.

The Household Questionnaire had 16 sections that were filled out by the interviewer during the first visit to the household, and in the following four weeks according to the following scheme:

FIRST VISIT

1. Initial visit
 - A. List of household members
 - B. Summary of presence in the household
 - C. Information on migration
 - D. Food consumption during the last 7 days

WEEK 1:

2. Education and literacy
3. Housing

WEEK 2:

4. Household economic activities
 - A. Land ownership
 - B. Production of crops including fruits and vegetables etc.
 - C. Cost of cultivation of crops (including fruits and vegetables)
 - D. Hypothetical questions on rental and sales market
 - E. Inputs and outputs of livestock raising activities
 - F. Input and outputs from fish cultivation and fishing/trapping of aquatic products
 - G. Inputs and outputs from forestry and hunting
 - H. Inventory of household non-agricultural economic activities during the past 12 months
5. Household liabilities
6. Household income from other sources, and other expenditures (partial non-food recall)

WEEK 3:

7. Durable goods and other expenses
 - A. Durable goods
 - B. Other expenditures
8. Construction activities in the past 12 months
9. Nutrition
 - A. Rice consumption
 - B. Other food
 - C. Vulnerability
10. Fertility and child care
 - A. Fertility history
 - B. Child feeding and vaccinations
11. Mortality

WEEK 4:

12. Health check of children
13. Current economic activity
 - A. Activity status during the past 7 days
 - B. Main occupations during the past 7 days
 - C. Employment and earnings in the last 12 months
14. Health
 - A. Illness during the past 4 weeks
 - B. Smoking information
15. HIV/AIDS
16. Victimization
 - A. Household security
 - B. Victim of theft
 - C. Victim of accidents

Once the month ended, the team went back to the NIS headquarters in Phnom Penh

Questionnaires from the same PSU was delivered to the Data Management team by the supervisor in a packet including all of the documents used and produced in the fieldwork, including maps, enumeration lists, questionnaires, diaries, etc.

Before going to the villages teams were briefed and introduced to minor adjustments of the interviewing procedure that had to be made as a result of monitoring activities and feed-back from the data processing.

ANNEX 5 contains an example (the first survey month) from the allocation of teams to PSUs.

3.4 Monitoring

Any survey of the CSES dimensions needs a comprehensive system for quality management and monitoring. Only then can deviations from the target be tended to in time to avoid shortfalls.

The CSES management group within NIS therefore set up a meticulous monitoring scheme to be implemented from the very beginning. The monitoring team did include at least five NIS staff. Commonly the DG of NIS has spent one week monthly while other

top ranked NIS officers have been out for two weeks on average. At times other officials from NIS or the Ministry have participated.

Inspections entailed both announced and unannounced visits. Every team was visited at least twice during their fieldwork periods. The purposes of these visits were several. One important purpose was to get a disciplinary effect on supervisors and enumerators from their knowledge that such inspections must be expected throughout the fieldwork month, including also at the very end of the diary month. Also important was to give feedback and encouragement to fieldworkers and to complement training by advice and suggestions and to sort out any problem that had arisen in the course of fieldwork in the village. Another area of concern was to ensure that the household listing and sampling was done in accordance with the procedures that were devised.

In *ANNEX 6* an example (January 2004) of the Field Supervision Plans is included, the other months look very much the same.

4 Data processing

4.1 First assessment

A first assessment of the ICT situation at the NIS was made in early 2003. The findings made clear that extra efforts were needed to cope with the vast volumes of data to be expected. Recommendations included an upgrade of equipment, a single LAN for data editing and entry operations, staff training, and the setting up of a data management system. Since UNDP funds were not deemed sufficient, NIS was encouraged to seek support elsewhere. It was at this stage, in August 2003, that the World Bank started to explicitly contribute to the survey by recruiting IT consultants for the work to be done.

4.2 Data management

Since NIS staffs were familiar with the CSPro software, which was used in the CSES 1999 and the Population Census 1998, it was decided to develop the data management system in CSPro. It was hoped that much of the previous data processing programs could be reused and thus save time and resources. However, it was soon found that the new set of questionnaires and the sheer increase in data volume made it necessary to start more or less from scratch.

The data management system had to deal with an average of 1,000 households a month and had to meet the following main criteria:

- 1 Ensure that the workflow would not cumulate any backlogs. That meant to make the data from fieldwork conducted in month-x electronically available by the second week of month x+2.
- 2 To minimize data entry errors and allow for the highest quality transcription of the information gathered in the field.
- 3 To provide good quality data to the survey analysts on a timely basis

The framework for the CSES data processing using CSPro was outlined in July/August 2003 with WB expert assistance. At that time the questionnaires were not finished and

the existing code tables in dire need of revision. The data management organization needed re-vitalization.

In November 2003 a first version of the data management system was developed and the main parts were tested and integrated i.e. the *data management* component, the *data editing* component, and the *data entry* component. At this time component manuals were also written.

The data management team was organized in the following way: *the Data Management Director* as the overall responsible person, and the *Data Management Chief* responsible for the daily operations. A *quality control team* was appointed as well as a *data entry team*.

The data management setup was based on a total of 24 workstations. It included three dedicated computers: one for the scanning of time-use sheets, one for data entry of the village questionnaire, and one for the editing and coding staff. For data entry of the household questionnaire, 18 workstations were allocated. These allowed for comfortably entering and correcting mistakes for an average of 60 PSUs per month, with an expected monthly productivity of 3.3 PSUs per operator, or equivalently an average of 2.3 completed households per day and per operator.

Two scanners were received in October seemed and while one was not working properly it was soon replaced.

To be able to provide backstopping via Internet an emailing service with a local provider was rented, providing at most ten mailboxes to be used by the staff.

The *Data Management Manual* is a 22 page document available at NIS. Only the Table of Content and the first chapter containing general guidelines are attached as *ANNEX 7*.

Data management and processing issues are fully covered in Canales (2003a), (2003b), and (2004).

4.3 Training

Training was provided by the World Bank on an intermittent basis. The first major effort was made in November 2003. In this initial training the manuals were the focal point. The contents were at the same time adjusted to solve the problems and difficulties found in the hands-on training directed at the key staff.

4.4 Data editing and coding

During the first months of data entry there were delays in the editing and coding operations that induced backlogs in the data entry operation due to the staff situation. In July 2004 a number of entry operators were retrained to assist in carrying out the coding work to eliminate the backlogs. A WB consultant was assigned to help out with problems from the coding operation and to correct possible errors in the entry program.

The *Coding and Editing Manual* is a 47 page document available at NIS. Only the Table of Content and the first chapter containing general guidelines are attached as *ANNEX 8*.

4.5 Data entry

The data entry module was developed in CSPro. During the testing phase up to December 2003 the survey a number of bugs or errors were identified and corrected. The module was thereafter continuously improved in the following months to enhance speed as well as quality based on the analyses of data already entered into data files

A separate module for *automated capture of time-use sheets* by using scanners was independently provided by an external consultant firm. The module was integrated into the data management system and extensively tested. One operator was appointed for scanning, and trained in the use of the module.

The *Data Entry Manual* is a 15 page document available at NIS. Only the Table of Content and the first chapter containing general guidelines are attached as *ANNEX 9*.

4.6 Reference database

In July 2004 a new set of CSPro solutions for data export was developed. It was then possible to deliver the CSES data in different formats for statistical tabulations and analysis using e.g. SPSS or STATA.

In order to secure the CSES data and facilitate for the production of statistical tables and other forms of dissemination, a *reference database* in the SQL standard format was designed and set-up. To populate the reference database, preformatted SPSS data files have been used. This database will constitute the *reference point* for the CSES 2004 to which all further analysis will refer.

The main goal when designing the database was to set up a system for easy retrieval of micro-data for different users and groups. An important matter, when it comes to all databases optimized for retrieval, is the need for good and reliable metadata that systematically describes all variables for the survey. Thus a big effort was made to create metadata for the more than 500 variables included in the CSES reference database. Another important factor dealing with metadata is to complement the database with standard classifications and other types of codes. Therefore, separate code tables were created and joined together with underlying data to form data views.

After the first version of the reference database was established, all data were checked for inconsistencies compared to their source data in SPSS. All errors were corrected and changed directly in the database.

To further facilitate for data retrieval a metadata driven Visual Basic application for extraction of CSES variables was designed and implemented. In its present state the extraction application works only with data stored in the SQL Server. When all data are finally checked and corrected the SQL Server database could easily be migrated to MS Access and copied to a CD-Rom together with an extraction tool and distributed outside the NIS. Extracted data are stored as local text files and can be exported to any software that supports import of external data e.g. Excel, SPSS, STATA or SAS. They can also be exported into client/data applications e.g. CAMInfo or PC-Axis.

5 Tabulation and dissemination

A first set of CSES statistical tables was suggested in connection with the analysis of the pilot survey in August 2003. During the survey, workshops have been held to present intermediate statistics for selected subject matter areas. In June 2005 a workshop was arranged in Phnom Penh for dissemination and discussion of eight major subject matter reports to national and international stakeholders. These reports have since been edited into one volume, the *Summary Subject Matter Report*.

A dissemination plan was simultaneously laid out with expert assistance. This plan includes a recommendation to produce a booklet with highlights from the CSES results. A selection of indicators from the CSES should be added to the CAMInfo CD. There should also be separate CDs with macrodata to cater for more professional users' needs to get more detailed data and to make further analysis. Also tailor-made statistics should be offered and produced on request. General CSES information and as much data as possible should also be put on the NIS website not least to serve international users. However, the capacity of the website is currently very limited, which makes it difficult to use as an efficient dissemination tool without upgrading of server capacity and additional resources. The Data User Service Centre should also get adequate training in understanding the CSES to be able to give professional assistance to users.

Generally the statistical awareness in the Cambodian society is limited. The users of statistics in Cambodia are fairly few and seem to have little experience in demanding and using statistics. There is a strong need for training and capacity building of users. Workshops are therefore recommended in connection with the release of future CSES subject matter reports. These should also generate press releases. To work actively and strategically with media is also a recommendation.

References

Statistics Sweden/UNDP reports

- Bäcklund S., Söderberg L. Planning the data processing for the Cambodia Socio-Economic Survey. (2003).
- Dalén J. Measurement problems in poverty estimation in Cambodia. (2005)
- Davidsson G., Pettersson H. Survey Design for the Cambodia Socio-Economic Survey (2003)
- Henningsson B., Lindgren E. Trainer training for interviewing purposes. (2003)
- Johansson S. Designing the Next Socioeconomic Survey for Cambodia. (2003a)
- Johansson S. Designing the Next Socioeconomic Survey for Cambodia (2). (2003b)
- Johansson S. Designing the Next Socioeconomic Survey for Cambodia (3). (2003c)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (1). (2004a)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (2). (2004b)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (3). (2004c)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (4). (2004d)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (5). (2004e)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (6). (2004f)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (7). (2005a)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (8). (2005b)
- Johansson S. Conducting the 2003/04 Cambodia Socioeconomic Survey (CSES) (9). (2005c)
- Johansson S. Analyzing the 2003/04 Cambodia Socioeconomic Survey (1). (2005d)
- Johansson S. Analyzing the 2003/04 Cambodia Socioeconomic Survey (2). (2005e)
- Neupert R. Demographic estimates and revised population projections. (2005).
- Pettersson H. Calibration of sampling weights for the Cambodian socio-economic surveys. (2004)
- Pettersson H. Calibration of sampling weights for the Cambodian socio-economic surveys (2). (2005)
- Söderberg L. Establishing the CSES reference database. (2005)
- Vorwerk P. Analysis of the pilot survey for the Cambodia Socio-Economic Survey. (2004)
- Östergren A., Åström L. Dissemination issues for the CSES 2004. (2005)
- Technical Report on Survey Design and Implementation. (2005)

The above reports are electronically available at the National Institute of Statistics, UNDP, and Statistics Sweden as well as relevant memos not listed here.

The World Bank

- Canales V. Mission Report (1). (2003a)
- Canales V. Mission Report (2). (2003b)
- Canales V. Mission Report (3). (2004a)

Annex 1. NIS Survey Staff and Expert Advisors

SURVEY PERSONNEL

I. NIS Core Group and Supporting Staff (10 Persons)

1	H.E. San Sy Than	Survey Director
2	Mr. Seng Soeurn	Deputy Survey Director
3	Mr. Khieu Sary	Field work Coordinator
4	Mr. Saint Lundy	Data Processing Coordinator Sampling Coordinator
5	Mr. Mich Kanthul	Questionnaire Design Coordinator
6	Mr. Tith Vong	Admin. officer
7	Mr. Vy Heang	Accountant
8	Mr. Lim Penh	Repairman of Computer network
9	Mr. Sek Sarin	Computer maintenance and cleaning staff
10	Mr. Sar Sem	

II. NIS and MOP Field Staff 125 Persons (Group I)

1	Mr. Bin Chanthea	43	Mr. Kim Net	85	Mr. Pho Sokha
2	Mr. Bin Gneann Rith	44	Mr. Koe Chantra	86	Mr. Phok Visal
3	Mr. Bou Noch	45	Mr. Kong Phalla	87	Mr. Pon Sophal
4	Mr. Bouy Someathea	46	Mr. Kong Seng	88	Mr. Pov Sam Ol
5	Mr. Bun Tha	47	Ms. Ty Chankagna	89	Mr. Roeurn Sothea
6	Mr. Chan Naroun	48	Ms. Ky Boreth	90	Ms. Ros Bundane
7	Mr. Chan Nipol	49	Mr. Ky Long	91	Mr. Ros Punlork
8	Mr. Chan Samath	50	Mr. Leng Vansak	92	Mr. Ros The
9	Ms. Chay Dalen	51	Mr. Lenh Heang	93	Mr. Ros Vireythun
10	Ms. Chay Dala	52	Mr. Lmot Samkol	94	Mr. Sa Chivan
11	Mr. Chea Eng	53	Mr. Louk Samphis	95	Mr. Sam Samroeurn
12	Mr. Chea Ngeth	54	Mr. Ly Srey Sothy	96	Mr. Sar Puthevy
13	Mr. Chea Veasna	55	Mr. Mak Huch	97	Mr. Sarin Prasidh
14	Mr. Cheav Vathna	56	Mr. Mao Bo	98	Mr. Seang Rith
15	Mr. Chhay Sokha	57	Mr. Mao Saron	99	Mr. Sem Dara
16	Mr. Chheam Kimim	58	Mr. Meas ChanTavee	100	Mr. Sier Ketya Pea
17	Ms. Chhin Phearum	59	Mr. Mech SoKhorn	101	Mr. Sim Ly
18	Mr. Chhom Pisith	60	Mr. Men Nirintivorn	102	Ms. Sisovat Satya
19	Mr. Chhun Chim	61	Mr. Moeurng Rado	103	Mr. So Sokhim

20	Mr. Chim Sayoth	62	Mr. Mok Seng	104	Mr. So Tonnere Mr. Sok
21	Ms. Chum Rumnear	63	Mr. Mom Sathya	105	Sovannarong
22	Ms. Chuon Sereyrath	64	Mr. Nan Sothara	106	Mr. Som Bony
23	Mr. Dom Chanveasna	65	Ms. Net Sophy	107	Ms. Som Manith
24	Mr. Doung Channara Mr. Duch Cham	66	Mr. Net Visal	108	Ms. Som Marinath
25	Roeun	67	Mr. Nguon Nor	109	Mr. Song Theary
26	Mr. Eam Hour	68	Mr. Nhem Solyvan	110	Mr. Tan Kantol
27	Ms. Eng Sophoan	69	Ms. Nhem Srey	111	Mr. Thav Se
28	Mr. Ham Bora	70	Mr. Nim Sitha	112	Mr. Tieng Ran
29	Mr. Han Danin	71	Mr. Nop Heang	113	Mr. Tim Phy Yary
30	Mr. Heang Sovithyea	72	Mr. Nou Phyrun	114	Mr. Tuch Minear
31	Mr. Heng Daravuthy	73	Mr. Noun Nisay Kosal	115	Mr. Uch Soeurn
32	Mr. Hok Thy	74	Mr. Nuth Chea	116	Mr. Un Chamroeun
33	Mr. Hor Darith	75	Mr. Om Deth	117	Mr. Uy Savan
34	Mr. Hor Sarin	76	Mr. Ou Tavin	118	Mr. Van Khan
35	Mr. Hout Vitou	77	Mr. Ouk Eam	119	Mr. Vat Sophan
36	Mr. Hun Visal	78	Ms. Ouk Morokat	120	Mr. Vong Sina
37	Mr. Huy Kangon	79	Ms. Ouk Samasaty	121	Mr. Yan Kong
38	Mr. Hy Kimkry	80	Ms. Ouk Synan	122	Mr. Yann Chamnith
39	Mr. Khem Khy	81	Mr. Ouk Ty	123	Mr. Yi Sophal
40	Mr. Khiev Pagna	82	Mr. Ouk Vichet	124	Mr. Yim Saonith
41	Mr. Khin Song	83	Mr. Phan Chinda	125	Mr. Yim Sothea
42	Mr. Khun Kim Sroe	84	Mr. Phay Pisith		

III. Provincial, NIS and MOP Field Staff 125 Persons (Group II)

1	Mr. Ang At	43	Mr. Moeurng Sam	85	Mr. Sek Chhunleng
2	Mr. Chan Sokun	44	Mr. Moul Sokun	86	Mr. Sek Somonn
3	Mr. Chan Tara	45	Mr. Moul Vanak Mr. Moung	87	Mr. Seng Bunroeurn
4	Mr. Chay Satha	46	Sopheap	88	Mr. Seng Kear
5	Mr. Chea Sothy	47	Mr. Neang Saroeun	89	Mr. Seng Lay
6	Mr. Chea Sovann	48	Ms. Ngoun Channa	90	Mr. Seng Serey Roth
7	Mr. Chhet Chhay	49	Mr. Nhek Vuthy	91	Ms. Sim Navy Mr. Sin
8	Mr. Chhim Samoeun	50	Ms. Nhoung Kunthea	92	Chansopheak
9	Mr. Chhin Chantha	51	Mr. Nong Pang	93	Mr. So Vannak
10	Mr. Chin Dina	52	Mr. Noun Silout	94	Ms. Sok Chanthet
11	Ms. Chorn Phally	53	Mr. Nuon Sithy	95	Mr. Som Sariem
12	Mr. Chorn Saphon	54	Mr. Nuth Nysaiy	96	Mr. Song Nit
13	Mr. Choun Chamroeun	55	Mr. Oeur Sophal	97	Mr. Souk Soeurn
14	Mr. Chum Oeurn	56	Mr. Ong Narrieng	98	Mr. Soun Vannar

15	Mr. Dor ChanKour	57	Mr. Ou Sarin	99	Mr. Sous Say
16	Mr. Eat Vanna	58	Mr. Ouch Seng	100	Mr. Sum Neang
17	Mr. Eng Nareth	59	Mr. Ouk Sarem	101	Ms. Suy Lanich
18	Mr. Hee Sross	60	Mr. Pech Buntha	102	Mr. Tai Yong
19	Mr. Heng Chhenghour	61	Mr. Pen Sopha	103	Mr. Tan Sopheak
20	Mr. Hok Phearun	62	Mr. Peng Kiengsay	104	Mr. Teing Vantha
21	Mr. Hou Savat	63	Mr. Pha Engsry	105	Mr. Tep Saroeun
22	Ms. I Kimsorth	64	Mr. Phach Chesda	106	Mr. Tep Veasna
23	Mr. Im Nhen	65	Mr. Phan Sokha	107	Mr. Tes Puoth
24	Mr. Im Sokleng	66	Mr. Phauk Chantha	108	Mr. Theng Cheang
25	Ms. Im Sothea	67	Ms. Phoung Vichany	109	Mr. Thoeur Thok
26	Mr. Ka Den	68	Mr. Pok Chiv	110	Mr. Thuy Ly
27	Ms. Khem Sopheap	69	Ms. Pol Sophea	111	Mr. Tiev Choulong
28	Mr. Khim Sophearun	70	Mr. Pov Setha	112	Mr. Tim Bunthan
29	Ms. Koh Andeth	71	Mr. Prok Saboeun	113	Mr. Tin Bunthoeun
30	Mr. Kong Monny	72	Mr. Prum Bunnary	114	Mr. Yim Saonith
31	Mr. Kong Phalla	73	Mr. So Sokhim	115	Ms. Tuy Sunary
32	Mr. Kong Savath	74	Mr. Reath Bunthoeurn	116	Mr. Um Ry
33	Mr. Loy Tep	75	Mr. Ros Sarith	117	Mr. Vann Sann
34	Mr. Mak Phearum	76	Mr. Ros Sary	118	Mr. Ven Phoan
35	Mr. Man Pheareak	77	Mr. Sam Kosal	119	Mr. Yi Dany
36	Mr. Mao Chhorn	78	Mr. Sam Lin	120	Mr. Yi Rith
37	Mr. Mao Sirann	79	Mr. Sam Rethchantharo	121	Mr. Yi Touch
38	Mr. Men Mach	80	Mr. Samrith Samon	122	Mr. Yim Kosal
39	Mr. Men Rith	81	Mr. San Sophat	123	Mr. Yim Oeurn
40	Mr. Min Thy	82	Ms. Sao Kimhy	124	Mr. Yim Rath
41	Mr. Min Chhan	83	Mr. Sao Prumvuthy	125	Mr. Yin Sophal
42	Mr. Moeur Kao	84	Mr. Seak Pengkeang		

IV. Data Editors Staff (22 Persons)

1	Ms. Hang Ly Na	9	Mr. Teav Rongsa	16	Ms. Mak Chanary
2	Ms. Em Samoeurn	10	Mr. Yem Sopharum	17	Ms. Tho Sam Chin
3	Ms. Tong Chhayrine	11	Ms. Hout Karolin	18	Ms. Long Forsevy
4	Mr. Vann Suon	12	Ms. Chhun Chhavy	19	Mr. Khin Sovorlak
5	Mr. Kong Chinda	13	Ms. Orn Davin	20	Ms. Heng Vichet
6	Mr. They Kheam	14	Mr. Ouch Monysetha	21	Mr. Khin Bunna
7	Mr. Pich Pothy	15	Mr. Chhoun Sothy	22	Mr. Kem Sidony
8	Mr. Oukchay Panhara				

V. Data Processing Staff (22 Persons)

1	Mr. Yip Thavvin	9	Mr. Sum Vannsan	16	Mr. Vy Sovyl
---	-----------------	---	-----------------	----	--------------

2	Mr. Lay Sophat	10	Mr. Mao Chhem	17	Ms. Van Camarat
3	Ms. Mam Manith	11	Mr. Nim Saomony	18	Ms. Hong Dany
4	Ms. Khon Nary	12	Ms. Khieu Madry	19	Ms. Krem Somaly
5	Mr. Khieu Khemarin	13	Ms. Oun Len	20	Mr. Chhun Bonarith
6	Ms. Kong Sreyny	14	Ms. San Sopha	21	Ms. Rin Sitha
7	Ms. Mao Vannoeurn	15	Ms. Mom Seila	22	Ms. Peng Napy
8	Ms. Mey Sokhantey				

Statistics Sweden

Professor Sten Johansson	Chief Advisor
Mr. Sten Bäcklund	Statistician, Project Coordinator
Mr. Ronnie Andersson	Statistician, ICO Director
Dr. Jörgen Dalén	Statistician, Index Theory
Ms. Gunilla Davidsson	Questionnaire Design
Mr. Henrik Gustafsson	Statistician
Ms. Birgit Henningsson	Interviewer Training
Ms. Eva Lindgren	Interviewer Training
Mr. Lars Lundgren	Statistician, Time Use
Dr. Ricardo Neupert	Demographer
Ms. Berit Olsson	Publishing
Dr. Hans Pettersson	Statistician, Sampling
Mr. L-J Söderberg	IT
Dr. Peter Vorwerk	Statistician
Ms. Cecilia Westström	Publishing
Ms. Lena Åström	Dissemination
Ms. Anna Östergren	Dissemination

The World Bank

Mr. Victor Canalez	Data Processing
Ms. Chorching Goh	WB Coordinator
Ms. Beatriz Godoy	Questionnaire Design, Enumerator Training

Mr. Paul Glewwe	General Advisor
Ms. Claudia Marchena	Data Processing
Mr. Juan Munoz	General Advisor

Annex 2. Sampling design and estimation

Estimation of variances

A total is estimated from the sample by the estimator

$$\hat{Y} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}$$

where

$$w_{hij} = \frac{A_h}{n_h A_{hi}} \cdot \frac{M_{hi}}{m_{hi}}$$

A_h = number of households in stratum h according to census

A_{hi} = number of households in PSU i in stratum h according to census

M_{hi} = number of households in PSU i in stratum h according to survey listing

m_{hi} = number of households in the sample from PSU i in stratum h

n_h = number of PSU: s selected in stratum h

The precision in an estimate of a total \hat{Y} is a function of the variance

$$Var(\hat{Y}) = \sum_{h=1}^L \frac{n_h}{n_h - 1} \left[\sum_{i=1}^{n_h} (y'_{hi})^2 - (\sum y'_{hi} / n_h)^2 \right]$$

where

$$y'_{hi} = \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}$$

A ratio, mean or proportion is estimated by:

$$\hat{R} = \frac{\hat{y}}{\hat{x}}$$

With the variance:

$$Var(\hat{R}) = Var(\hat{Y}) + \hat{R}^2 Var(\hat{X}) - 2\hat{R} Cov(\hat{Y}, \hat{X})$$

Estimation

Under this design the sampling weight for a selected household k in selected enumeration area hij is:

$$w_{hijk} = \frac{M_h}{n_h \cdot M_{hi}} \cdot \frac{E_{hi}}{1} \cdot \frac{M_{hij}}{m_{hij}} \dots\dots\dots(1)$$

Where

M_h = number of households in stratum h according to census

M_{hi} = number of households in village hi according to census

n_h = number of villages selected

E_{hi} = number of enumeration areas in village hi

M_{hji} = number of households in EA hij according to listing by interviewer.

m_{hji} = number of households interviewed in EA hij

Some of the EA:s selected for the survey have boundaries that are difficult to identify in the field. In such cases there is a risk that the interviewer makes mistakes by wrongly including households outside the EA or excluding households within the EA. One would hope that these mistakes even out in the long run but there are reasons to believe that the cases of wrong exclusion outweigh the cases of wrong inclusion. The result will be under-listing of households and estimates of totals that are biased downwards. A way of avoiding this is to calculate the weights by:

$$w_{hijk} = \frac{M_h}{n_h \cdot M_{hi}} \cdot \frac{M_{hi}^*}{m_{hij}} \dots\dots\dots(2)$$

Where M_{hi}^* = Number of households in village hi according to the village chairman

An advantage with the alternative weights (2) as compared to the design weights (1) is that the alternative weights adjust for any under-listing of households in the EA or segment. The alternative weights will also show less random variation than the design weights. One should note that the weights according to (2) rely completely on the village chairman's statement of the total number of households and not on a full listing of the households in the village. There are certainly "random" errors in the chairmen's statements and there may be a downward bias due to out-of-date household registers kept by the chairmen. However, this will be a minor problem because the weights will be calibrated so that estimates of total households from the sample will agree with census projections of the total number of households. It was decided to use method (2) to calculate the sampling weights.

Calculation of sampling weights

The calculation of the sampling weights is shown in a step-by-step manner in the file

sample weights for 900 villages.xls . There is also a copy of the file in SPSS-format. Some of the villages are very large. The best procedure would have been to put the very large villages (villages with a size M_{hi} larger than M_h/n_h) in a separate stratum. This was not done. As a result there are a few villages where the inclusion probability exceeds 1.00 and consequently the first stage sampling weight is below 1.00. To rectify this we set the first stage sampling weight (WI) equal to 1.00 for these villages. When doing so we had to adjust the weights downwards for the other villages in the stratum in order to have the same sum of weights for the stratum as before the adjustment. (The original first stage weights are called *prelWI* in the Excel-file and the adjusted weights are called WI . The corrections are calculated in the file: correction of weights for big villages.xls).

The second stage sampling weights are calculated as M_{hi}^*/m_{hij} , i.e. the number of households in the village (according to the chairman) over the number of sampled households (10 or 20). There were actually two variables indicating the number of households in the village, one was the number obtained by the interviewer (variable *E_HHs*) and the other was the number obtained by the supervisor (*HHs_Vill*). A check revealed some apparent data entry errors, which were corrected. The second stage weights are shown as variable $W2$ in the excel file.

The household sampling weights (W_{prel}) are calculated by multiplying WI by $W2$. All the sampled households in the village get the same household weight. A check of the weights revealed that there were a few extremely low and high weights. These “outliers” will tend to inflate the variance for some estimates. We decided to trim the weights by adjusting the extreme weights towards the center. Weights above 300 were adjusted downwards to 300 and weights below 30 were adjusted upwards to 30. In all only six weights were adjusted, the set of weights after trimming is named W in the file. The distribution of the household weights is shown in diagrams 1 and 2 below. The variation of the weights reflects changes in village sizes between the Census 1998 and the time of the survey. If the current number of households were the same as during the census in all the sample villages there would be no variation at all in the weights. The rather large variation in the weights is by and large a consequence of the long time lag between the census and the survey.

The distribution of the weights for the urban households shows a tendency towards bimodality. Most of the weights are centered around 100 – 120 but there is a cluster of weights around 200 – 220. The reason for this slight abnormality is that the proportional allocation of the sample to strata was not strictly followed. In two urban strata the sample size was below proportion, resulting in substantially larger sampling weights.

Diagram 1: Distribution of sampling weights for urban households

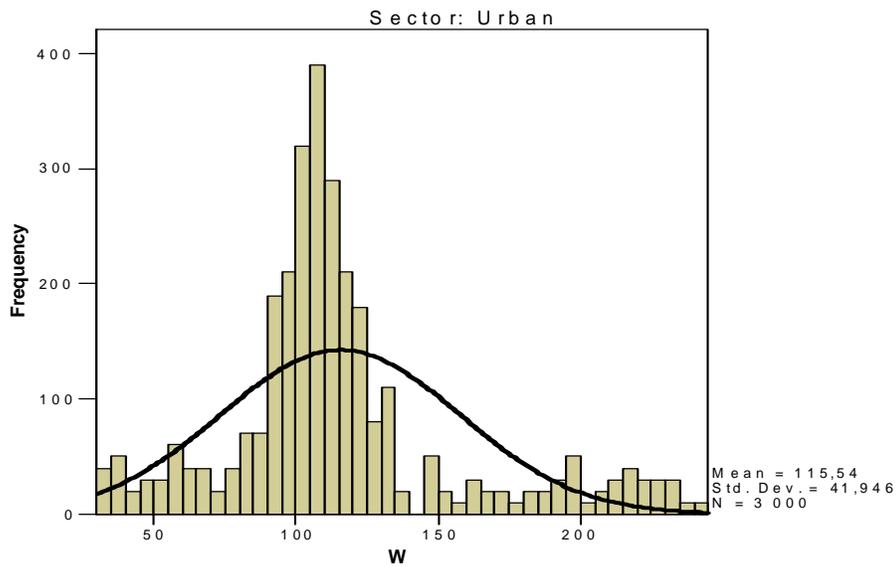


Diagram 2: Distribution of sampling weights for rural households

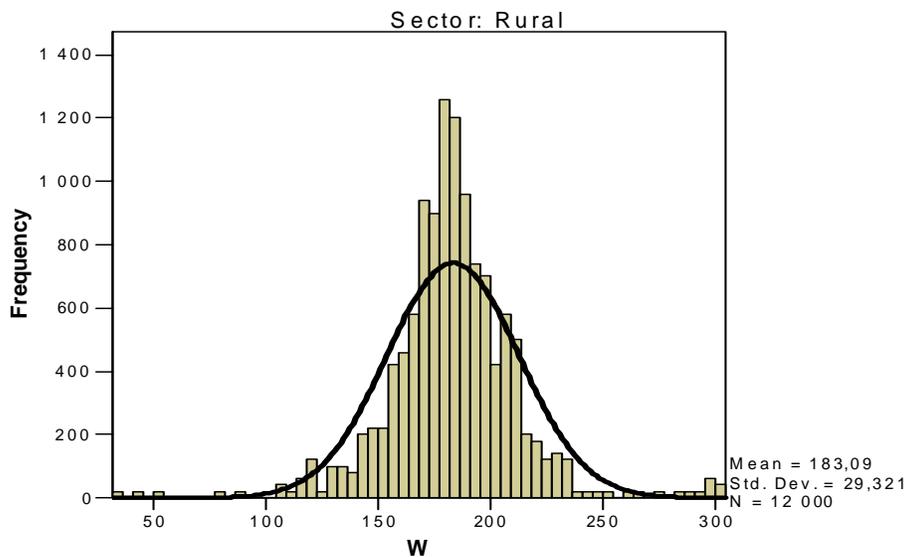


Table 1 below shows population and household estimates using the sampling weights we have calculated and 95% confidence intervals for the estimates. For comparison we include the estimates from the CIPS2004 survey.

The CIPS survey used two questionnaires, Form A and Form B, to collect information about number of households and population in the selected enumeration areas. Form A was filled in during the listing of all the households in the enumeration area. Form B was completed only for the randomly selected households in the enumeration area (30 households/EA). Estimates of total population and number of households from Form A-data are thus based on a larger sample than estimates from Form B-data. The confidence intervals will consequently be wider for the Form B estimates.

Table 1: Estimates of total population and total number of households (Mill.) from CSES and CIPS. (CIPS-A and CIPS_B are estimates based on Form A and Form B respectively)

Survey	Estimate	Estimate	95% conf. interval	
			lower	upper
CSES	Number of households (Mill.)	2.544	2.504	- 2.584
CIPS-A	Number of households (Mill.)	2.530	2.485	- 2.579
CIPS-B	Number of households (Mill.)	2.530	2.491	- 2.569
CSES	Total population (Mill.)	12.623	12.443	- 12.805
CIPS-A	Total population (Mill.)	12.824	12.572	- 13.075
CIPS-B	Total population (Mill.)	12.346	12.098	- 12.594
CSES	Average household size	4.97	4.92	- 5.01
CIPS-A	Average household size	5.07	5.03	- 5.10
CIPS-B	Average household size	4.88	4.84	- 4.92

The three estimates of number of households are close. The CIPS-A and CIPS-B estimates are identical; this is what we should expect from the design. The estimates of total population differ somewhat; the difference between the two CIPS estimates is fairly large. A part of the difference could be attributed to sampling error but even considering the sampling error it seems that the listing procedures in Form A and Form B give different results. (The data on number of persons per household may be better established in Form B where each person in the household is listed during the interview?)

We also get differences in the average household size in CIPS-A and B as a result of the differing population estimates from Form A and B.

The estimates from CSES are in-between the two CIPS estimates for total population and average household size, which is somewhat reassuring. However, all three estimates of total population are substantially below the total population for 2004 calculated by demographic projections based on census data. This will be discussed in the next section.

Annex 3. Sample allocation to strata

Province	Number of villages			Sample of 900 villages		
	Urban	Rural	Total	Urban	Rural	Total
01 Banteay	59	544	603	17	30	47
02 Battambang	62	507	569	24	39	63
03 Kampong Cham	31	1717	1748	8	98	106
04 Kampong	26	520	546	7	24	31
05 Kampong Speu	56	1263	1319	7	35	42
06 Kampong Thum	55	682	737	12	31	43
07 Kampot	15	462	477	6	32	38
08 Kandal	20	1067	1087	9	63	72
09 Kaoh Kong	11	116	127	5	6	11
10 Kracheh	74	183	257	14	11	25
11 Mondol Kiri	14	84	98	1	1	2
12 Phnom Penh	405	232	637	90	25	115
13 Preah Vihear	31	173	204	4	5	9
14 Prey Veng	42	1094	1136	10	59	69
15 Pursat	63	412	475	10	19	29
16 Rattanak Kiri	16	224	240	3	4	7
17 Siem Reab	77	805	882	19	34	53
18 Sihanouk Vile	85	0	85	26	0	26
19 Stueng Treng	17	111	128	4	3	7
20 Svay Rieng	18	672	690	4	30	34
21 Takeo	40	1076	1116	7	48	55
22 Oddar Meanchey	21	83	104	4	3	7
23 Krong Keb	16	0	16	5	0	5
24 Krong Pailin	58	0	58	4	0	4
	1312	12027	13339	300	600	900

DISTRIBUTION OF SAMPLE VILLAGES BY PROVINCE MONTH AND TEAM (MONTH 1)

PSU No.	P_C	P_N	D_C	D_N	C_C	C_N	V_C	V_N	No._EAs	S_EA	U / R	T_HHs	HHS_S	Workload	Month	Tea
01036	01	Banteay Mean Chey	06	Serei Saophoan	08	Tuek Thla	01	Keab	5	4	1	659	10	W008	01	0
01037	01	Banteay Mean Chey	06	Serei Saophoan	08	Tuek Thla	03	Tuek Thla	6	2	1	796	10	W008	01	0
01042	01	Banteay Mean Chey	08	Svaay Cheek	01	Phkoam	05	Oun	1	1	2	69	20	W008	01	0
02001	02	Bat Dambang	01	Banan	02	Kantueu Pir	05	Kampang Kaeut	2	2	2	120	20	W031	01	0
02054	02	Bat Dambang	06	Moung Ruessei	11	Preaek Chik	03	Chhkae Kham Praeus	2	1	2	214	20	W031	01	0
02002	02	Bat Dambang	01	Banan	04	Chheu Teal	10	Anlong Taa Mei	3	1	2	304	20	W053	01	0
02055	02	Bat Dambang	07	Rotonak Mondol	02	Andaek Haeb	02	Taa Saanh	3	2	2	206	20	W053	01	0
03030	03	Kampong Cham	06	Kampong Siem	03	Kien Chrey	06	Kien Chrey Knong	2	1	2	136	20	W070	01	0
03031	03	Kampong Cham	06	Kampong Siem	05	Kaoh Mitt	03	Kaoh Paen Ko	2	1	2	172	20	W070	01	0
03058	03	Kampong Cham	11	Ou Reang Ov	01	Ampil Ta Pok	02	Prey Sralau	3	1	2	164	20	W085	01	0
03059	03	Kampong Cham	11	Ou Reang Ov	01	Ampil Taa Pok	11	Pou Svaay Ming	2	1	2	146	20	W085	01	0
03089	03	Kampong Cham	15	Stueng Trang	08	Preah Andoung	01	Preaek Sdei	3	3	2	327	20	W100	01	1
03090	03	Kampong Cham	15	Stueng Trang	10	Preaek Kak	08	Preah	2	1	2	271	20	W100	01	1
04022	04	Kampong Chhnang	06	Rolea B'ier	05	Kouk Banteay	07	Kang Meas	2	2	2	197	20	W117	01	1
04023	04	Kampong Chhnang	06	Rolea B'ier	08	Prasneb	06	Prasneb	2	1	2	173	20	W117	01	1
05020	05	Kampong Speu	06	Phnum Sruoch	05	Krang Dei Vaay	01	Banteay Rokaa	1	1	2	78	20	W135	01	1
05021	05	Kampong Speu	06	Phnum Sruoch	07	Ou	09	Ou Lvea	1	1	2	80	20	W135	01	1
06016	06	Kampong Thum	03	Stueng Saen	01	Damrei Choan Khla	01	Balang Khang Kaeut	3	2	1	239	10	W150	01	1
06017	06	Kampong Thum	03	Stueng Saen	01	Damrei Choan Khla	02	Balang Khang Lech	5	2	1	430	10	W150	01	1
06018	06	Kampong Thum	03	Stueng Saen	01	Damrei Choan Khla	02	Balang Khang Lech	5	3	1	430	10	W150	01	1
06019	06	Kampong Thum	03	Stueng Saen	01	Damrei Choan Khla	03	Pou Bak Ka	3	2	1	251	10	W150	01	1

Annex 4. Team allocation

06028	06	Kampong Thum	04	Praasaat Balangk	01	Doung	03	Kruor	1	1	2	102	20	W153	01	1
06029	06	Kampong Thum	04	Prasat Balangk	03	Phan Nheum	06	Prohut	2	2	2	150	20	W153	01	1
07015	07	Kampot	03	Chhuk	10	Mean Chey	03	Tul	2	2	2	229	20	W168	01	2
07016	07	Kampot	03	Chhuk	13	Trapeang Bei	03	Trapeang Bei	3	1	2	309	20	W168	01	2
08014	08	Kandal	02	Kien Svaay	10	Preaek Thmei	02	Kaoh Krabei	9	1	2	867	20	W186	01	2
08015	08	Kandal	02	Kien Svaay	11	Samraong Thum	03	Chrouy Dang	6	4	2	609	20	W186	01	2
08045	08	Kandal	08	Angk Snuol	07	Krang Mkaak	12	Prambei Mum	1	1	2	44	20	W202	01	2
08046	08	Kandal	08	Angk Snuol	11	Peuk	03	Trapeang Peuk	1	1	2	83	20	W202	01	2
19003	19	Stueng Traeng	01	Seesaan	07	Taa Laat	01	Svaay Rieng	2	2	2	164	20	W344	01	2
19004	19	Stueng Traeng	03	Siem Paang	04	Srae Sambour	04	Peam Khes	1	1	2	79	20	W344	01	2
12006	12	Phnom Penh	01	Chamkar Mon	01	Tonle Basak	09	Center 9	3	2	1	386	10	W228	01	2
12007	12	Phnom Penh	01	Chamkar Mon	01	Tonle Basak	10	Center 10	6	6	1	585	10	W228	01	2
12008	12	Phnom Penh	01	Chamkar Mon	01	Tonle Basak	15	Center 15	15	3	1	1 417	10	W228	01	2
12009	12	Phnom Penh	01	Chamkar Mon	02	Boeng Keeng Kang Muc	01	Center 1	2	1	1	173	10	W228	01	2
12066	12	Phnom Penh	04	Tuol Kouk	01	Phsar Depou Muoy	01	Center 1	3	1	1	264	10	W243	01	3
12067	12	Phnom Penh	04	Tuol Kouk	01	Phsar Depou Muoy	07	Center 7	2	1	1	131	10	W243	01	3
12068	12	Phnom Penh	04	Tuol Kouk	01	Phsar Depou Muoy	08	Center 8	1	1	1	119	10	W243	01	3
12069	12	Phnom Penh	04	Tuol Kouk	03	Phsar Depou Bei	02	Center 2	2	1	1	143	10	W243	01	3
12111	12	Phnom Penh	07	Ruessei Kaev	06	Ruessei Kaev	02	Phum Samma Ki	11	4	2	1 062	20	W259	01	3
12112	12	Phnom Penh	07	Ruessei Kaev	07	Tuek Thlaa	03	Phum Sleng Roleung	7	1	2	588	20	W259	01	3
14021	14	Prey Veang	04	Kanhchriech	01	Chong Ampil	06	Prey Svaay	3	1	2	261	20	W277	01	3
14022	14	Prey Veang	04	Kanhchriech	03	Kdoeang Reay	17	Boeng Anhchaanh	1	1	2	56	20	W277	01	3

14055	14	Prey Veang	10	Prey Veang	09	Samraong	09	Prohut	2	1	2	156	20	W295	01
14056	14	Prey Veang	10	Prey Veang	11	Tuek Thlaa	15	Traok	1	1	2	99	20	W295	01
16001	16	Rattanak Kiri	01	Andoung Meas	03	Nhaang	09	Peeng	1	1	2	29	20	W311	01
16005	16	Rattanak Kiri	02	Ban Lung	02	Labansiek	01	Phum Muoy	7	5	1	1 028	10	W311	01
16006	16	Rattanak Kiri	02	Ban Lung	03	Yeak Laom	01	Lon	1	1	1	85	10	W311	01
17003	17	Siem Reab	01	Angkor Chum	07	Taa Saom	06	Tumpung	1	1	2	104	20	W315	01
17053	17	Siem Reab	14	Vaarin	02	Lvea Krang	02	Ou Tey	2	1	2	134	20	W315	01
10003	10	Kratie	01	Chhloung	08	Preaek Saamaan	01	Chheu Teal Phluoh	4	2	2	656	20	W219	01
10018	10	Kratie	03	Preaek Prasab	02	Chrouy Banteay	04	Khsaach Tob	3	1	2	237	20	W219	01
20032	20	Svay Rieng	07	Svaay Teab	08	Romeang Thkaol	06	Kranhung	3	1	2	278	20	W360	01
20033	20	Svay Rieng	07	Svay Teab	09	Sambuor	07	Sambuor	2	2	2	204	20	W360	01
21009	21	Takaev	02	Bati	13	Tnaot	06	Pey	2	1	2	185	20	W366	01
21010	21	Takaev	02	Baati	14	Trapeang Krasang	08	Kou Duol	2	1	2	115	20	W366	01
24001	24	Krong Pailin	01	Khand Pailin	03	Toul Lvea	05	O Chra Leu	1	1	1	31	10	W342	01
24002	24	Krong Pailin	02	Sala Krau	03	Stueng Kach	02	Ta Ngen Kraom	1	1	1	69	10	W342	01
24003	24	Krong Pailin	01	Pailin	01	Sangkat Pailin	06	Voat	1	1	1	140	10	W342	01
24004	24	Krong Pailin	02	Sala Krau	03	Stueng Kach	03	Ou Beng	1	1	1	74	10	W342	01

Annex 5. Field supervision plan

	Province of Departure	Date of Departure	Province of Arrival	Date of Arrival
	Phnom Penh	18/01/2004	Kandal	18/01/2004
	Kandal	20/01/2004	Siem Reab	20/01/2004
	Siem Reab	25/01/2004	Phnom Penh	25/01/2004
	Phnom Penh	15/01/2004	Kampong Cham	15/01/2004
	Kampong Cham	20/01/2004	Prey Veng	20/01/2004
	Prey Veng	22/01/2004	Svay Rieng	22/01/2004
	Svay Rieng	26/01/2004	Kampong Thum	26/01/2004
	Kampong Thum	29/01/2004	Phnom Penh	29/01/2004
	Phnom Penh	15/01/2004	Kampong Speu	15/01/2004
	Kampong Speu	19/01/2004	Takeo	19/01/2004

	Takeo	24/01/2004	Kampot	24/01/2004
	Kampot	29/01/2004	Phnom Penh	29/01/2004
	Phnom Penh	16/01/2004	Kandal	16/01/2004
	Kandal	20/01/2004	Siem Reab	20/01/2004
	Siem Reab	26/01/2004	Oddor Meanchey	26/01/2004
	Oddor Meanchey	30/01/2004	Phnom Penh	30/01/2004
	Phnom Penh	15/01/2004	Pursat	15/01/2004
	Pursat	18/01/2004	Battambang	18/01/2004
	Battambang	21/01/2004	Banteay Meanchey	21/01/2004
	Banteay Meanchey	25/01/2004	Phnom Penh	25/01/2004
	Phnom Penh	16/01/2004	Kandal	16/01/2004
	Kandal	20/01/2004	Siem Reab	20/01/2004
	Siem Reab	26/01/2004	Oddor Meanchey	26/01/2004
	Oddor Meanchey	30/01/2004	Phnom Penh	30/01/2004
	Phnom Penh	15/01/2004	Pursat	15/01/2004
	Pursat	18/01/2004	Battambang	18/01/2004
	Battambang	21/01/2004	Banteay Meanchey	21/01/2004
	Banteay Meanchey	25/01/2004	Phnom Penh	25/01/2004

Annex 6. Excerpt from the data management manual

Table of Contents

CHAPTER 1. THE CSES AND THE DATA MANAGEMENT COMPONENT

FIELDWORK OVERVIEW

DATA MANAGEMENT OVERVIEW

THE LIFE CYCLE OF ONE MONTH OF FIELD WORKLOAD

- 1 dm: Data Management Manual
- 2 de: Data Entry Operator's Manual
- 3 ed: Coding and Editing Manual

CHAPTER 2. ORGANIZATION OF THE DATA MANAGEMENT

FRAMEWORK FOR THE DATA MANAGEMENT SUPPORT THE MODEL OF THE DATA
MANAGEMENT ORGANIZATION

THE LIFE CYCLE OF ONE MONTH OF FIELD WORKLOAD

- 4 Daily operation management
- 5 Workflow model
- 6 Qualifications of the staff

CONTROL CATALOGUE

- 1 Control catalogue module: mastering the workflow
- 2 How to measure the progress

DATA FILES NAMING CONVENTION

QUESTIONNAIRES TAILORED FOR DATA MANAGEMENT PURPOSES

AUTOMATED CAPTURE OF TIME USE SHEETS

EVALUATING THE QUALITY OF THE FIELDWORK: QUALITY CONTROL TABLES

CODING TABLES

- 1 Occupation's code
- 2 Industry code
- 3 Crops code
- 4 Cause of death
- 5 Diary expenditures
- 6 Consumption units
- 7 <not implemented> food calories
- 8 <not implemented> automated tool for the management of coding tables

CODES FOR FIELDWORK AND OFFICE STAFF

- 1 Fieldwork teams and fieldwork staff codes
- 2 Codes for the data entry operators
- 3 Quality control personnel

CHAPTER 3. COMPONENTS OF THE DATA MANAGEMENT SYSTEM

THE DATA ENTRY PROGRAM FOR HOUSEHOLDS (HHENTRY)

- 1 Missing values
- 2 Automatically skipped portions
- 3 Interpretation of messages
- 4 Partial save
- 5 Further considerations

THE CONSISTENCY PROGRAM (HHCHECK)

CHECKING DATABASE INTEGRITY, PARAMETRIC SELECTION TOOL (SELECT)

THE DATA ENTRY PROGRAM FOR THE VILLAGE QUESTIONNAIRE (VIENTRY)

OTHER SUPPLEMENTARY TOOLS

ADVISABLE FEATURES TO BE IMPLEMENTED BY THE DATA MANAGEMENT TEAM

The CSES and the data management component

The CSES will visit a very large sample of 15,000 households over 15 months. The data management component must therefore deal with an average of 1,000 households a month. The data management component should be able to meet the following main objectives:

- 1 *Ensure that the workflow will not cumulate any backlogs. In other words, make the datafiles of fieldwork conducted in month- x available by the second week of month $x+2$.*
- 2 *To minimize data entry errors and allow for the highest quality transcription of the information gathered in the field.*
- 3 *To provide good quality databases to the survey analysis on a timely basis.*

Fieldwork overview

There are 50 teams allocated to the fieldwork. Each month there will be 25 teams working at the field, with a workload covering 2, 3 or 4 sample points or PSUs. The fieldwork plan has been scheduled in order to gather around 60 households monthly per team.

Teams are headed by a **Team Supervisor** and integrated by four **Interviewers**.

For a given month the team moves to the PSU one week before in order to achieve preparatory tasks (enumerate the households and select those to be interviewed). Then, each of selected households is submitted to the first part of the questionnaire, where the household composition is established. For the whole month of operation, the household fills up a diary of expenditures and incomes, and the interviewer visits it repeatedly to complete the remaining portions of the questionnaire.

Once the month ended, the team gets back to the NIS central headquarters in Phnom Penh, where the Supervisor delivers the product of the job to the Data Management team. Each of PSUs is delivered in a packet including all of the documents used and produced in the fieldwork, including maps, enumeration lists, questionnaires, diaries, etc. In this manual, this is referred to as the PSU-packet, and it is the production for the whole workflow.

Data management overview

The Data Management component must enter the information to data files with the maximum quality possible. To meet this objective, there are two separated teams:

- 3 **Data Entry** room: working with around 20 Data Entry Operators for the household questionnaires, plus other staff to enter the village questionnaires and to capture the information of time-use sheets, this group works with smart data entry programs which control the accuracy and reliability of the entry job. The Data Entry room is headed by a **Data Entry Supervisor**.
- 4 **Quality Control** team: integrated by around 5-6 Editors, this group is in charge of receiving and checking the contents of PSU-packets, coding all the needed fields, and prepares the PSU-packets for the data entry job. Once the PSU-packet entered, they receive the error reports, analyze all the possible errors and mistakes, and determine the corrective actions to be carried out. This team is headed by a **Quality Control Supervisor**.

The activity of both teams above mentioned is coordinated by a **Data Management Supervisor**, who works continuously with the Supervisors.

time, the **Data Management Supervisor** should have the authority to allocate more staff to this task.

All this requires a tight control and careful daily evaluation on the workflow situation to adopt the best measures leading to ensure the final delivery date (the 12th day of the month $x+2$) will always be respected.

Documentation for the data management component

There are three main manuals on the data management system. All are in standard word' document files, and have names starting with the prefix CSES, followed by two letters referring to the specific manual, ending with two digits referring to the version of the document. These manuals and their residence files are as explained below.

1 *dm*: the Data Management Manual

This document, the **Data Management Manual**, is stored in files named *CSESdmxx.doc*.

2 *de*: Data Entry Operator's Manual

For the **Data Entry Operator's Manual**, look at the file *CSESdexx.doc*.

3 *ed*: Coding and Editing Manual

The **Coding and Editing Manual** is placed into files *CSESexx.doc*.

To print his manual, it is also required to print some of coding tables non embedded in the document. All of the related coding tables are also included in the accompanying excel sheets described later in the Chapter 2 of this document.

Each version is date-stamped in the cover and footers of the documents, and also in the *properties* of the file.

Although only the last version of each manual should be used at any moment, it is useful to keep track of former versions.

Newer versions are to be periodically issued in response to changes, enhancements and improvements on the elements of the system.

Annex 7. Excerpt from the data editing and coding manual

Table of Contents

CHAPTER 1. GENERAL GUIDELINES FOR DATA EDITING

HOW THE FIELDWORK IS ORGANIZED
BASIC INSTRUCTIONS
HOW THE WORKFLOW IS ORGANIZED AT THE OFFICE

CHAPTER 2. RECEPTION AND CODING

RECEIVING THE QUESTIONNAIRES FROM THE FIELDWORK
CHECKING THE ENTIRE PSU PACKET
CODING OPEN QUESTIONS IN THE HOUSEHOLD QUESTIONNAIRE
CODING TABLE: CROPS
CODING TABLE: CAUSE OF DEATH
CODING TABLE: OCCUPATION
CODING TABLE: INDUSTRY
CODING TABLE: EXPENDITURE ITEMS IN THE DIARY
CODING TABLE: MEASUREMENT UNITS OF THE EXPENDITURES IN THE DIARY

CHAPTER 3. EDITING ERRORS

GENERAL POLICY ON EDITING
HOW TO DEAL WITH SKIP ERRORS
FORMAT OF MESSAGES
MESSAGE DESCRIPTION AND ACTIONS TO BE TAKEN

- 7 Data file management
- 8 Household cover
- 9 Section 01:List of household members
- 10 Section 02: Household cover
- 11 Section 03: Housing
- 12 Section 04: Household economic activities
- 13 Section 05: Household liabilities
- 14 Section 06: Household income from other sources
- 15 Section 07: Durable goods and other expenditure
- 16 Section 08: construction
- 17 Section 09: Nutrition
- 18 Section 10: Fertility and child care
- 19 Section 11:Mortality
- 20 Section 12: Health check of children
- 21 Section 13: Current economic activity
- 22 Section 14: Health
- 23 Section 15: HIV/AIDS
- 24 Section 16: Victimization
- 25 Diary cover
- 26 Diary expenditures
- 27 Diary

incomes

APPENDIX 1. TOLERANCE LIMITS FOR WEIGHT AND WEIGHT OF CHILDREN

APPENDIX 2. CODES FOR CROPS

APPENDIX 3. CODES FOR CAUSE OF DEATHS

APPENDIX 4. CODES FOR OCCUPATION

APPENDIX 5. CODES FOR INDUSTRY

APPENDIX 6. CODES FOR EXPENDITURE ITEMS IN THE DIARY
APPENDIX 7. CODES FOR MEASUREMENT UNITS IN THE DIARY

General guidelines for data editing

This guidelines address to the Quality Control team, and provide instructions for the Editors working on, at one side, receiving and coding the questionnaires arrived from the field and, at the other side, working on editing and fixing the errors reported for the data files produced at the data entry workshop.

How the fieldwork is organized

There are 50 teams allocated to the fieldwork. Each month there will be 25 teams working at the field, with a workload covering 2, 3 or 4 sample points or PSUs. The fieldwork plan has been scheduled in order to gather around 60 households monthly per team.

Teams are headed by a **Team Supervisor** and integrated by four **Interviewers**.

For a given month the team moves to the PSU one week before in order to achieve preparatory tasks (enumerate the households and select those to be interviewed). Then, each of selected households is submitted to the first part of the questionnaire, where the household composition is established. For the whole month of operation, the household fills up a diary of expenditures and incomes, and the interviewer visits it repeatedly to complete the remaining portions of the questionnaire.

Once the month ended, the team gets back to the NIS central headquarters in Phnom Penh, where the Supervisor delivers the product of the job to the Data Management team. Each of PSUs is delivered in a packet including all of the documents used and produced in the fieldwork, including maps, enumeration lists, questionnaires, diaries, etc. In this manual, this is referred to as the PSU-packet, and it is the production for the whole workflow.

Basic instructions

Each PSU-packet is assigned by the **Quality Control Supervisor** exclusively to one of Editors. The appointed Editor is responsible for the PSU from its arrival to the office, up to the final approval of the information contained in the corresponding PSU-file.

Each member of the Quality Control staff has a code provided by the Quality Control Supervisor. This code must be written on the cover of each of questionnaires of the PSU. For all practical purposes, the code written is equivalent to the signature of the Editor in charge.

The Quality Control staff use red ballpoint pens to differentiate their annotations from the remaining staff participating in the survey.

Remind it is expected that fieldwork enumerators use blue pens, and fieldwork supervisors use blank pens. Data entry operators are expected to use green ballpoint pens.

How the workflow is organized at the office

The Data Management Manual of the CSES there is a detailed description on the model of the data management organization. Let's summarize here the key facts for the Quality Control team:

- 1 A fieldwork Team Supervisor returns to the office and delivers the PSU-packet (i.e. all of the questionnaires for a PSU where his team has worked for one month).
- 2 At this moment, one member of the Quality Control team is allocated to the

- Reception of the PSU-packet.
- 3 The PSU-packet is checked to ensure that all of the expected questionnaires are in place, along with the control sheets of the field work, maps, plans, and all of the auxiliary materials used in the fieldwork. Any differences are to be immediately discussed and analyzed by the Team Supervisor with the Quality Control Supervisor, who may refuse the reception of incomplete PSU-packets and return them to the Team Supervisor.
 - 4 Received PSU-packets are divided into three parts: TimeUse' sheets are sent to automatic scanning; Village questionnaires are transferred to the specialized operator working in his/her exclusive workstations; and the Household questionnaires (including Diary) must be fully coded and prepared for the data entry.
 - 5 Once the PSU-packet ready for data entry of the Households, the packet is delivered by the Quality Control Supervisor to the Data Entry Supervisor. The PSU is now said to be "in activity".
 - 6 The Data Entry Supervisor assigns the PSU-packet to a data entry operator, who must enter the questionnaires into a PSU-file. At the end of each data entry session, a consistency report is issued, displaying the errors or problems found in the information of the PSU-file. Errors reported may be solved or not by the operator with the support of the Data Entry Supervisor.
 - 7 Once all of the Household questionnaires of the PSU have been entered, the Data Entry Supervisor transfers the files of the PSU to his own workstation. The Data Entry Supervisor delivers the consistency report to the Quality Control Team along with the Household questionnaires.
 - 8 The consistency report is checked by the Editor in charge of the PSU, who may decide to correct some of the reported problems. Correcting instructions are returned to the Data Entry Supervisor, who makes the assigned operator to apply the correcting instructions.
 - 9 This correcting cycle may be repeated several times. Once the Editor in charge decides there is no more corrections to be made, the PSU becomes ready to be approved.
 - 10 Approval of a PSU is a solemn compromise sealed off by the Quality Control Supervisor and the Data Entry Supervisor, with the authorization of the Data Management Supervisor.
 - 11 Once the PSU approved, all of the PSU-files are transferred to the \CSES\Appro' folder of the concentrator workstation in charge of the Data Management Supervisor, and erased from both the Data Entry supervisor and the operator workstations. The PSU-packet is recomposed with all of the original questionnaires and documents, and placed in a separated storage. None of the approved materials (data files and packets) should never get back to the active' circuit.

The Data Management Supervisor types in the Control catalogue of the operation 6 key events in the life cycle of each PSU: (1) the beginning of the fieldwork, (2) the end of the field work, (3) the reception at the office, (4) the preparation of the PSU-packet, (5) the moment the PSU is assigned to a data entry operator, and (6) the final approval of the PSU.

Annex 8. Excerpt from the data entry manual

Table of Contents

CHAPTER 1. GENERAL GUIDELINES FOR DATA ENTRY

INTRODUCTION

QUALITY RATHER THAN SPEED

DEVIATING PATHS AND MISSING VALUES

SCREENS AND PAGES OF THE QUESTIONNAIRE

FIELD NAMING CONVENTION

MESSAGES TO THE OPERATOR

LAUNCHING A DATA ENTRY SESSION

WHAT THE OPERATORS SHOULD NEVER DO AND HOW THEY ARE SANCTIONED

CHAPTER 2. SPECIFIC INSTRUCTIONS

HOUSEHOLD COVER

SECTION 01: LIST OF HOUSEHOLD MEMBERS

SECTION 02: HOUSEHOLD COVER

SECTION 03: HOUSING

SECTION 04: HOUSEHOLD ECONOMIC ACTIVITIES

SECTION 05: HOUSEHOLD LIABILITIES

SECTION 06: HOUSEHOLD INCOME FROM OTHER SOURCES

SECTION 07: DURABLE GOODS AND OTHER EXPENDITURE

SECTION 08: CONSTRUCTION

SECTION 09: NUTRITION

SECTION 10: FERTILITY AND CHILD CARE

SECTION 11: MORTALITY

SECTION 12: HEALTH CHECK OF CHILDREN

SECTION 13: CURRENT ECONOMIC ACTIVITY

SECTION 14: HEALTH

SECTION 15: HIV/AIDS

SECTION 16: VICTIMIZATION

DIARY COVER

DIARY EXPENDITURES

DIARY

INCOMES

General guidelines for data entry

This manual is aimed at the Data Entry team of the CSES. It contains instructions for the Data Entry operators working at the data entry central facility to deal with the Household questionnaire and its integrated Diary of Expenditures and Incomes

Introduction

The data entry for the Cambodian Socio-Economic Survey 2003-2005 (CSES) will be carried out using the software package **CSPro**. The data entry program has been built to meet the following requirements:

- 1 to check for acceptable codes for questions
- 2 to follow skips and filters of the questionnaires
- 3 to check the consistency of the data as they are entered

Each PSU of the survey is recorded in a separated data file. Each PSU is assigned to a unique data entry operator, and recorded in a separated data file residing in the operator's workstation. The choice of isolated PSU-files instead of a one large, shared data file allows for greater levels of protection against power failures, hardware malfunctions or software misbehaviours.

The data entry operator enters the information of each household of the assigned PSU to the PSU-file. When all of the households are entered, the Data Entry Supervisor transfers the set of PSU-files to his own machine and assigns a new PSU to the operator.

Quality rather than speed

The data entry operator should keep in mind that this survey requires to pay careful attention to the accuracy of the job. This quality factor is rather more important than the speed.

The data entry program will not allow for entering values or codes that are out of defined ranges. It skips to the next question expected to be entered, according to the logic of the interview. The program will automatically move to another screen whenever the conditions of the interview ask for such an action, depending on logical decisions taken on the data already entered. Be aware that, if the operator enters an acceptable code or value, the program will react accordingly to the conditions specified in the questionnaire for that code, even if the entered code is not recorded in the interview. Consequently, data entry errors may lead to flows of data entry different from the flow of the actual interview. The only thing the system cannot do is to ensure that the operator is recording the same information provided in the questionnaire, so data entry operators must review and thoroughly check what they have typed in to ensure there are no mistakes.

As the program will control which question is the next to be entered, the operator must continuously watch at both the screen and the questionnaire, to check the program leads his/her activity to the same place as in the interview. Operators must not work "heads down", only reading the questionnaire: if the mistype any data field, the result may lead to major errors in the information recorded in the data files.

Deviating paths and missing values

Data entry activity will have sometimes to deal with two problems in entering the data from the questionnaires.

First, let's refer to *deviating paths*. If the operator finds some data in the questionnaire that is not requested by the data entry program, he/she should review all of the data previously entered. Likely the problem is due to a wrong value entered in one of preceding questions. The revision should check, field by field, that the data entered is the same as the response given in the questionnaire. The operator is expected to check, not just a few surrounding questions, but look at the whole page and to any previous pages possibly linked to the deviation.

If no mistakes are identified, the operator should alert the Data Entry Supervisor to perform a deeper revision of the exception and to identify and solve the problem.

Whether the Data Entry Supervisor is not able to locate the source of the problem, he will ask for assistance and advice to an Editor, or to the Quality Control Supervisor if necessary.

The second issue regards *missing values*. Sometimes the expected question was left

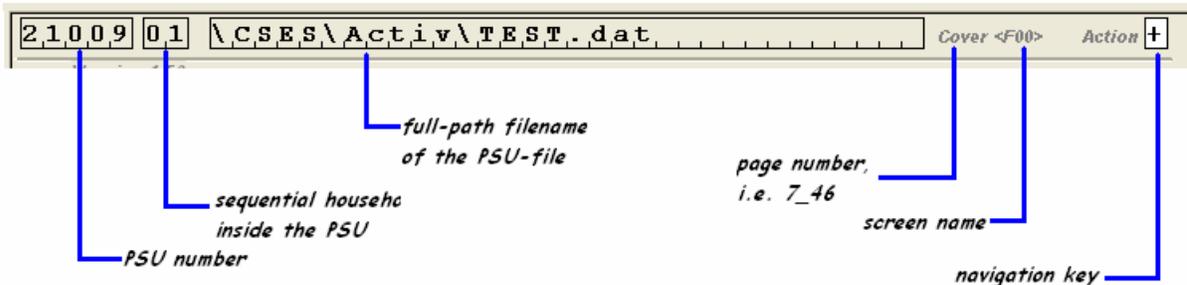
blank at the field. Whether it is an error of the interview, it is called a “missing” question, meaning that the enumerator simply did not put the question as expected. In these cases, following a through checking that the flow of the interview is the same as predicted by the program, data entry operator will fill out missing field with 9’s – 9 for a one-digit field, 99 for a 2-digit field, and so. Fields marked with missing values are later reported to the fieldwork supervisors, and they may lead to retraining or dismissing the involved staff. Any missing values are to be marked in the questionnaire using the green pens reserved to the data entry operators, to differentiate their annotations from the remarks made by other staff. Remind that fieldwork enumerators use blue pens, and fieldwork supervisors use blank pens. Coding and Editor staffs use red pens.

Screens and pages of the questionnaire

Each screen in the data entry program corresponds to one page of the Household questionnaire.

The first screen is a close reproduction of the cover of the questionnaire. It contains all of the identifying information, plus a time-stamped certificate produced by the program.

At the top of every screen there is the following information:



The operator does not need, excepting for taking a look on the households stored in the PSU-file, access to the standard left-pane tree provided by CSPro. Both the household-id and the PSU-filename are fully displayed the screens.

For this reason, the operator is always expected to work full-screen, without left-side pane, allowing for the widest display of the screen forms.

Field naming convention

The names of the data items in the program have the form *Qsspqqe*, where:

- 3 **Q**, the initial letter, refers to a question (navigation fields begin with a letter **Z**);
- 4 **ss** refers to the section number (i.e. 01 for the section 01, 12 for the section 12, etc). The cover of the household is referred to as **00**, while the diary cover is **40**;
- 5 **p** is the part of the section, if any. Some of sections have just one part (i.e. section 05) and this character is represented by an underscore. Other sections have several parts (i.e. section 04 has parts **A** thru **H**). and this character is the letter of the part;
- 6 **qq** is the question number, always with 2 digits;
- 7 **e** is the extension of the question number if needed. Some of questions have components named a/b/c or similar – in these cases, the data item ends with an uppercase letter corresponding to the component.

For example:

- 8 **Q02_03** is the name of question 3 in section 02 (*can the person write?*)
- 9 **Q04F09** is the name of question 9 in section 04 part F (*amount spent in fishing*)

activities)

10 **Q10A06A** is the name of question 6a in section 10 part A (*male children born to the respondent woman who are currently living in the household*)

During data entry sessions, the name of the data item where the cursor is placed is always displayed in the centre of the bottom line of the screen.

The operator should quickly relate the field name with its corresponding question and context in the questionnaire. It may be particularly useful for an efficient use of the **GoToField** operator command, whose shortcut is **F6**. The Data Entry Supervisor must ensure that all the operators know and are able to make fast use of the **GoToField** command.

Messages to the operator

The operator may receive warning and error messages when the program detects an abnormal condition or an inconsistency.

Messages are composed of a message-number, a qualifier and a descriptive text.

Messages are numbered with 6 digits as follows

sppqqx

where

ss is the section number (i.e. 02, 16)

p part of section if any. If the question has no parts, p is 0. Otherwise, part A is 1, B is 2, and so.

qq question number where the message is issued

x extension to the message

The qualifier is a letter telling of the level of the problem involved. Letter **E** stands for more serious errors, letter **W** is for warnings issued during data entry. In both cases, the data entry operator is expected to fix the problem by means of a complete checking of the data fields possibly involved in the reported conflict or even further across any of the preceding pages.

Most of messages having an **E** qualifier are initially designed to be solved at data entry time, and therefore should not appear in the consistency reports. However, the data entry program has been progressively modified to be more tolerant and to allow for the data entry operator to do not stop in front to these error messages.

Later consistency reports might include **E**-type messages, whose solution must be provided by Editors.

Launching a data entry session

The data entry program for an assigned PSU is to be launched by means the operator's menu. The only information to be provided is the PSU-number, a 5-digit number that identifies the PSU in a unique manner.

The operator's menu is an icon placed on the desktop of each data entry station. The Data Entry Supervisor will instruct each of his operators on the correct way to use this menu.

The menu has two options:

11 **create the PSU-file**: this option is used for the first time the operator works with the assigned PSU.

12 **modify the PSU-file**: for already existing PSU-files. It will allow for adding new

households, or for modifying the households previously entered in the PSU-file.

What the operators should never do and how they are sanctioned in case of

Due to some shortcomings of the current version of CSPro, the data entry program cannot disable a handful of very dangerous commands available to the operator. In consequence, the Data Entry Supervisor will ensure that no data entry operator makes use of the following commands:

13 *InsertGroupOcc*, or *F3*

14 *DeleteGroupOcc* or *F4*

15 *SortGroupOcc* or *F5*

16 *EditFieldNote* or *Ctrl+N*

Data entry operators are not allowed to look at the files, neither to edit them under any circumstances.

Data entry operators should never use the CSPro' designer to look at, or to edit, the entry program of this survey.

Operators will use their own menu to launch the CSES data entry program, and will refrain of any other mean of using the program.

The Data Entry Supervisor will immediately suspend any data entry operator who dares to violate one of these regulations.

The data files produced by the intruder operator are to be immediately deleted from all of the databases of the system, since its integrity cannot be ensured anymore. The workload made by him in the past or currently solved by him must be assigned to a different operator and restarted from scratch.

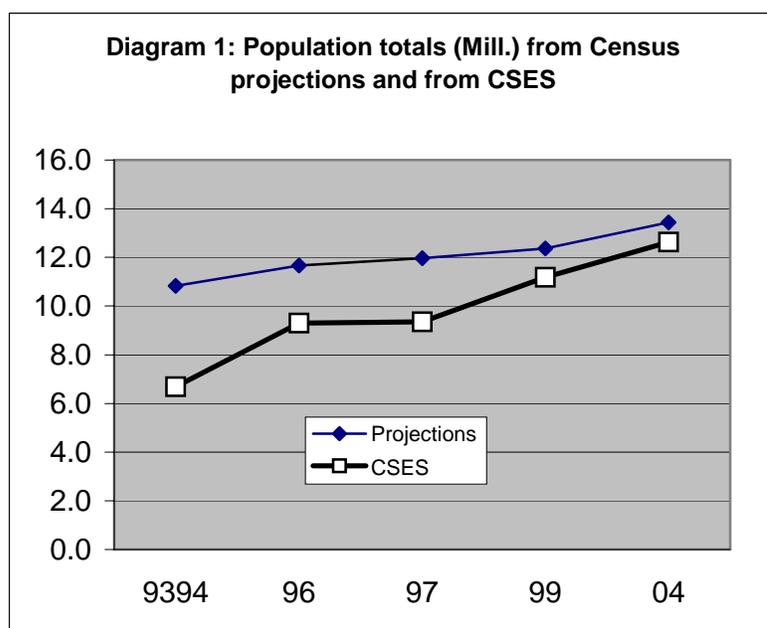
The Data Entry Supervisor will report to the Data Management Chief the suspended intruder operator. The Data Management Chief will ask to the high Direction of the NIS to separate forever the intruder from any data management activity.

The Data Entry Supervisor must ensure that all of his staffs know the above mentioned prohibitions.

Annex 9. Adjustment of the sampling weights for the five CSES surveys

The surveys

The first Cambodia Socioeconomic Surveys was carried out from the fourth quarter of 1993 through the third quarter of 1994 with a sample of 5 600 households divided into approximately equal-sized quarterly samples. The second CSES was done in two waves in May-July and October-December of 1996 with of a sample 9 000 households, the third in the last week of May and all of June in 1997. The fourth survey was done in two waves – in Jan-March and June-August of 1999 – with a sample of 6 000 households. All surveys have had to use truncated sampling frames to some extent. The “essential survey conditions” differ between the four previous surveys, there are differences regarding target population coverage, seasonal coverage and measurement methods. The estimates of population totals from the surveys reflect the different degrees of coverage in the surveys. The first three surveys did not cover all provinces; consequently these surveys will underestimate the population totals, as can be seen in diagram 1.



The large under-coverage in the earlier CSES surveys makes it necessary to adjust the sampling weights, i.e. to reweigh the survey data so that re-weighted estimates from the sample conform to known or projected household and population totals for key variables. Under-coverage is not a problem in the CSES2003/04 but adjustment of the design weights is still a good idea as it may improve the precision of the estimates for variables related to the adjustment variables. The weight adjustments will facilitate meaningful comparisons of estimates from the different CSES surveys

This kind of weighting adjustments is commonly applied in surveys to compensate for non response and non coverage, and to make weighted sample estimates conform to external values (Kalton and Flores-Cervantes 2003, Bethlehem and Keller 1987, Holt and Smith 1979)

Method

The sampling weights reflect the varying probabilities among the households to be included in the sample. The inclusion probabilities are defined by the sample design, the sampling weights are therefore often called “design weights”. When the sample design is implemented there will often be problems due to non-response and over- and/or under-coverage. The purpose of the calibration operation is to adjust the design weights so that the effects of non-response and coverage problems are reduced.

The adjustment of weights was done in two basic steps. The first step was to adjust the sampling weights (“design weights”) for households so that the estimates of total number of households from the surveys agree with the number of households according to census-based population projections. The second step was to use the adjusted household weights to estimate population totals by sex and 10 age groups for urban and rural areas in each of five zones. The estimates were compared to totals from the census projections and adjustment factors calculated for each sex and age group.

Population projections

Backward and forward projections of the population have been made based on the Population census 1998 (Neupert). The projections show the population in five-year age groups by sex for 15 provinces and 3 groups of provinces (by urban/rural). The projections cover the period 1993 – 2004. For our calculations the population was grouped into 10 age groups: 0-4, 5-9, 10-14,....., 35-39, 40-49, and 50+. The data are stored in the file *masterfile_population_projections.sav* . The projections by zone are shown in table 1.

Table 1: Population projections by zone ('000)

Zone	1993/94	1996	1997	1998	1999	2004
Phnom Penh	939	1 012	1 038	1 055	1 073	1 165
Plain	4 614	4 973	5 102	5 187	5 273	5 727
Tonle Sap	3 273	3 529	3 620	3 680	3 742	4 064
Coast	794	856	879	893	908	986
Plateau + mountain	1 205	1 299	1 332	1 354	1 377	1 496
Total	10 825	11 669	11 970	12 169	12 373	13 439

Projections of the number of households

The projections did not cover the number of households. The projected number of households in each province (by urban/rural) was therefore calculated by dividing the total population by the average household size for each survey year. The average household sizes per province were calculated from the five surveys and the Census.

The average household size in the Census was 5.22. This estimate is based on the census count where it is known that under-enumeration and age misreporting for the age group

0-4 is quite common. It is reasonable to assume that the under-numeration of children 0-4 will bias the population totals downward but will have little effect on household totals (because we can assume that the uncounted children in almost all cases live in households that were counted in the Census). The effect will be an underestimate of the household size.

The number of children 0-4 years old has been estimated using fertility and mortality estimates corresponding to the five years prior to the census date (obtained from the same census by indirect methods). The result of the application of this technique suggested that the number of males aged 0-4 years enumerated in the 1998 Census should be increased by a factor of 1.2713 and the number of females aged 0-4 should be increased by a factor of 1.2859 (Neupert).

If we apply these correction factors to the census data for children 0-4 and assume that these additional children are added to households already counted in the census we will get an adjusted household size of 5.42. The adjusted household sizes per province and urban/rural for the census year 1998 are shown in file: *hhsizes after adjustment for unaccounted children.xls*.

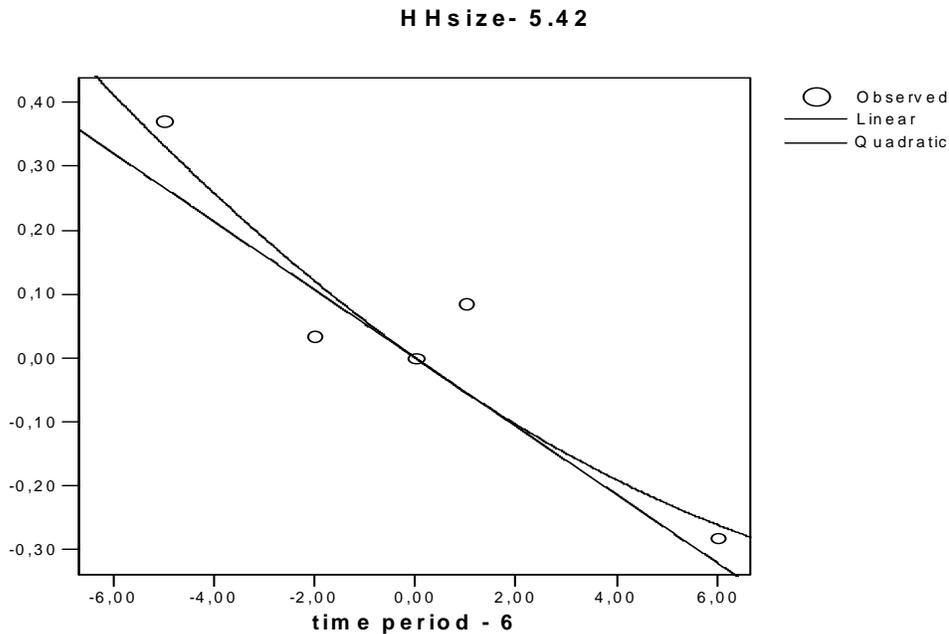
For the other years in the period 1993-2004 we only had estimates of household size from the surveys. For each of the five surveys we assumed that the rate of undercount of small children was the same as in the census. In each survey we estimated the number of under-enumerated children by applying the same factors as in the census adjustments above. The population totals from the surveys were adjusted upwards with the number of under-enumerated children. Estimated household sizes were calculated by dividing the adjusted population by the estimated number of households. The original and the adjusted estimates of household size are shown in table 3.

Table 2: Household sizes in the five CSES surveys.

	1993	1996	1997	1998	1999	2004
HH Size in the CSES surveys	5,60	5,27	4,95	5,23	5,34	5,00
Adjusted HH size in the CSES surveys	5,79	5,45	5,10	5,42	5,50	5,14
Quadratic trend, 3 obs	5,79	5,55	5,48	5,42	5,36	5,14
Quadratic trend, 5 observations,	5,75	5,54	5,48	5,42	5,36	5,16
Trend factor	1,0615	1,0223	1,0107	1,0000	0,9900	0,9518

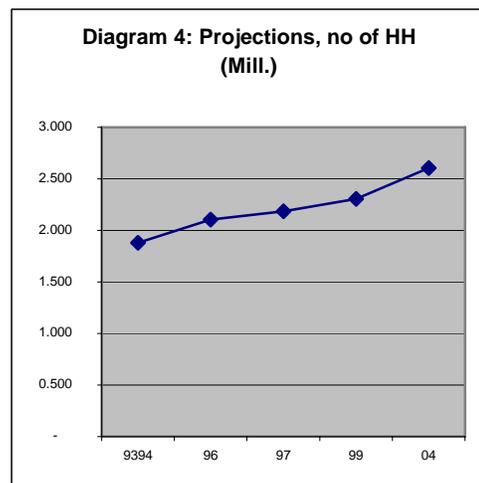
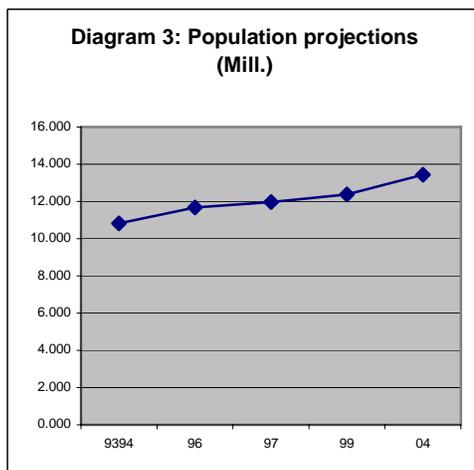
We used the adjusted estimates from the surveys and fitted a quadratic trend through the census value of 5.42. (The household size trend is shown in file *trend household sizes.xls*). In the first alternative we used only the first and the last value along with the census value (and consequently got a perfect fit). In the second alternative we used all observations except the 1997 value that was deemed of dubious quality. It was decided to use the second alternative for our calculations. The trend for the second alternative is shown in diagram 2.

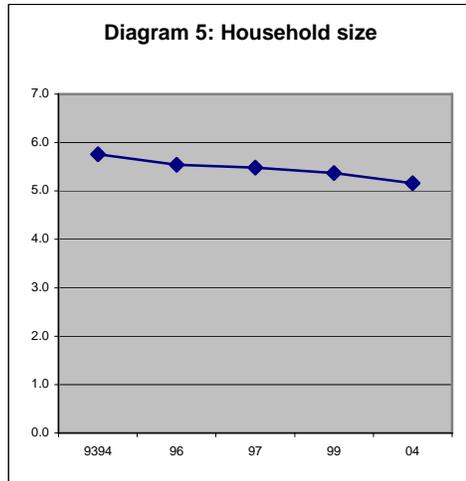
Diagram 2: Average household sizes from the surveys and fitted linear and quadratic trends



In order to get adjusted household sizes for each of the 18 province groups for urban and rural areas separately we applied the trend factors in table 1 to the estimated household sizes we had from the census.

The projected population totals were divided by the adjusted household sizes for each province group to get estimates of the total number of households. The number of households calculated from the population projections is shown in the file: masterfile_household_projections.sav. The projections on national level are shown in diagrams 3 – 5.





Adjustment of household sampling weights

For each survey we used the design weights to estimate the total number of households in each of the five zones (urban/rural). We divided the total number of households as calculated from the census projections by the estimates from the survey. These 10 ratios (adjustment factors, *wcorr*) were used to adjust the design weights. The design weights were multiplied by the adjustment factors to arrive at the adjusted weights. If, for example, the adjustment factor is 1.05 it means that the census projection in a particular group (zone*urban/rural) is 5% higher than the survey estimate. The design weights should therefore be adjusted upwards 5% in that group. The adjusted weights are named *finHHweights*. All the calculations are recorded in the excel file *preparation of adjusted weightsxx.xls* and in the SPSS syntax file *adjusting weightsxx.SPS*. The new weights are included in the SPSS files *householdfile xx.sav*

Differences in urban rural classification caused some problems. In CSES93/94 and CSES97 the whole of Phnom Penh was classified as urban while in the other surveys a part of Phnom Penh was classified as rural. In order to get the same urban classification we recoded the three districts in Phnom Penh from urban to rural in the CSES97. The new variable is called *urb2*. It was not possible to do the same reclassification in the CSES93/04 because no information on district code was available in the survey files.

The final household weights for CSES2004 are shown in the following diagrams.

Diagram 6: Distribution of household weights for urban households

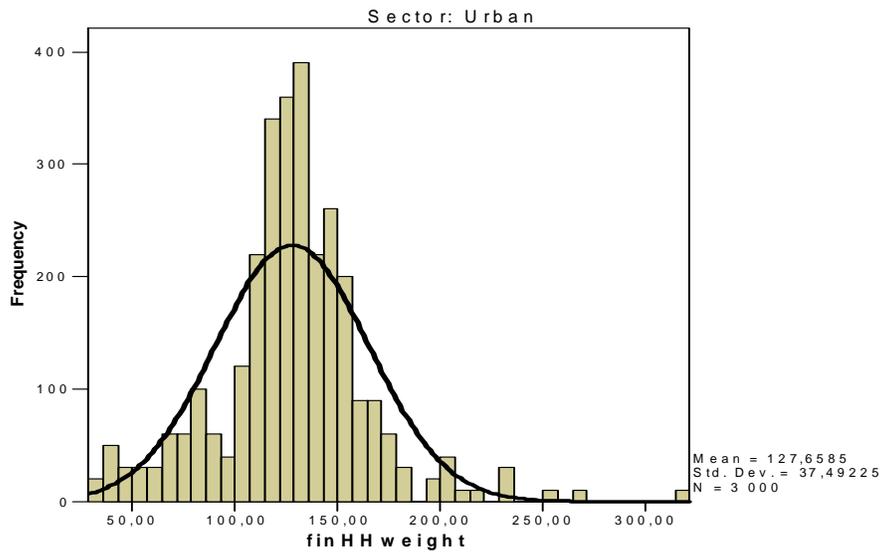
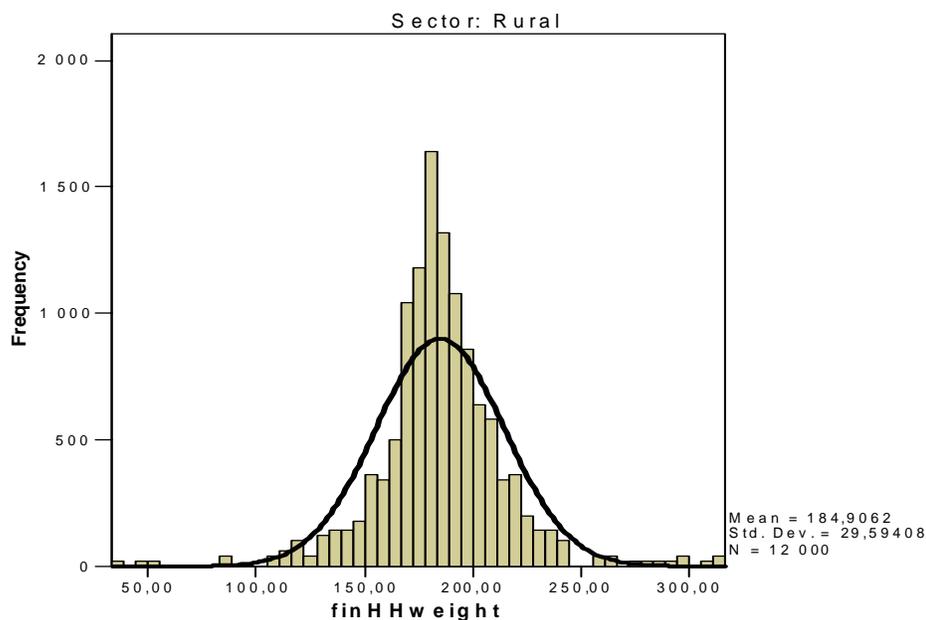


Diagram 7: Distribution of household weights for rural households



Adjustment of person sampling weights

From the surveys we estimated the total number of persons in each zone*urban/rural*sex*age-group (10 age-groups). In the estimation we used the new, adjusted household weights (*finHHweights*). The total numbers of people in the 200 groups according to the census projections were also calculated and divided by the totals from the survey. Ratios were calculated for each zone*urban/rural*sex*agegroup by dividing the census totals by the survey estimates. These ratios constitute the adjustment factors for the person weights (*pwcorr*). The *finHHweights* were multiplied by *pwcorr* to

give the *finPERweights*.

All the calculations are recorded in the excel file *preparation of adjusted weightsxx.xls* and in the SPSS syntax file *adjusting weightsxx.SPS* . The new weights are included in the SPSS files *personfile xx.sav* .

The sex*agegroup adjustment was made in each of the 900 sample villages. There are 20 sex*agegroups. Thus, the result of the weight adjustment is a set of 18 000 different person weights.

The final person weights for CSES2004 are shown in the diagrams below.

Diagram 8: Distribution of person weights for urban areas

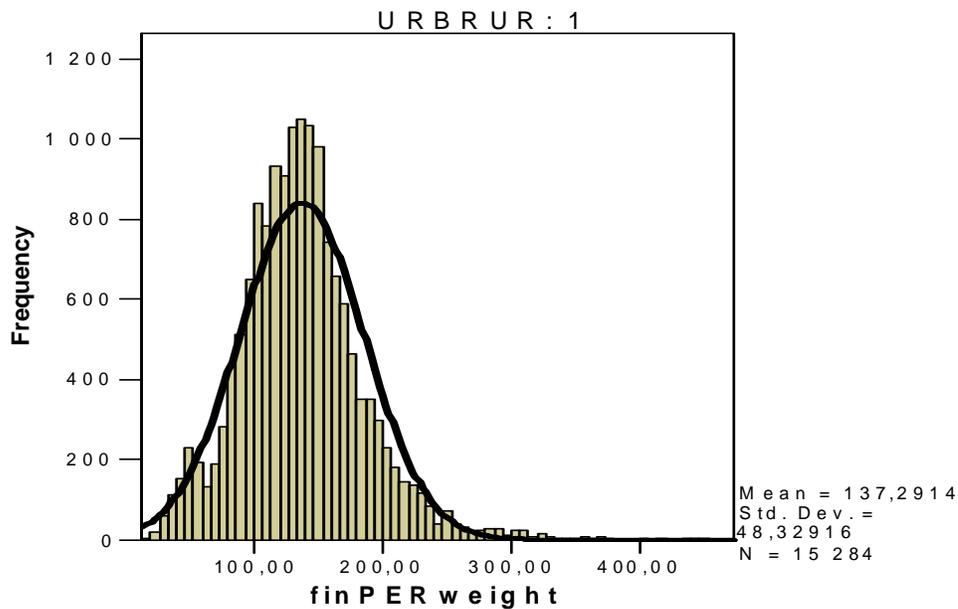
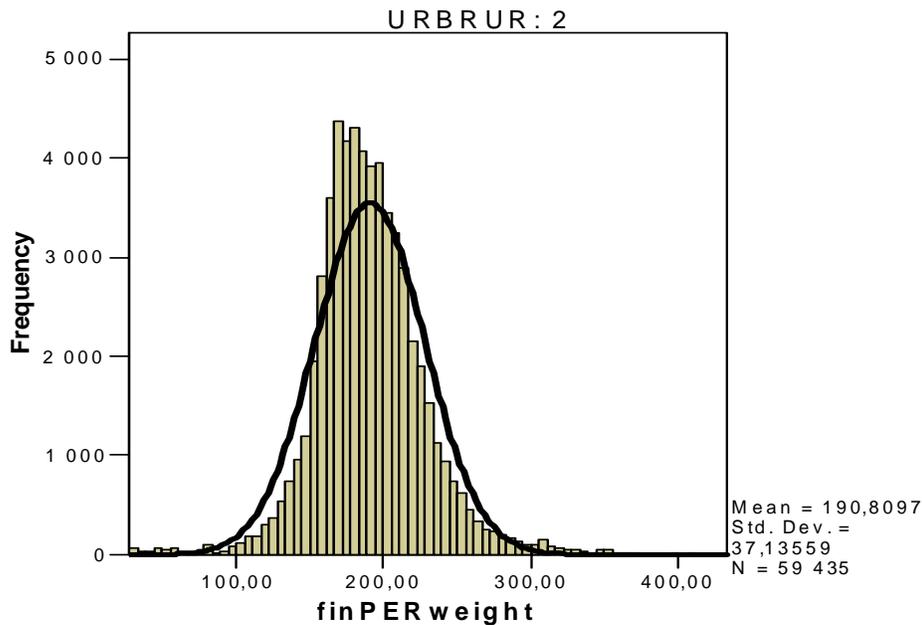


Diagram 9: Distribution of person weights for urban areas



Calendar year weights for CSES 2004

The CSES 2004 covers a 15-month period. For some purposes calendar year estimates are called for and in those cases we cannot use estimates based on the sampling weights for the full 900-village sample. There are several ways of solving this problem. We have opted for the alternative where the sampling weights for households not surveyed during the calendar year 2004 are set to zero and the remaining household weights are adjusted upwards to conform to the population and household totals. These adjusted weights are named *finHHweights12* and *finPERweights12* to separate them from the ordinary weights.

References

Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, Vol 19, No 2, pp. 81-97.

<http://www.jos.nu/Articles/abstract.asp?article=192081>

Bethlehem, J.G. and Keller, W.J.(1987). Linear Weighting of Sample Survey Data. *Journal of Official Statistics*, Vol 3, No2, pp.141-153

<http://www.jos.nu/Articles/abstract.asp?article=32141>

Holt, D. and Smith, T.M.F. (1979). Post Stratification. *Journal of the Royal Statistical*

Society, Series A, 142, 33-4
