By : Mr. Sok Kosal

KINGDOM OF CAMBODIA

Nation Religion King

SAMPLING TECHNICAL

CAMBODIA INTER-CENSAL POPULATION SURVEY 2004 (CIPS)

National Institute of Statistics, Ministry of Planning

Phnom Penh, Cambodia

Sponsored by :
United Nations Population Fund

October 2004

Sampling Design Report for
Cambodia Inter-Censal Population Survey (CIPS) 2004

I. Introduction

In some cases it may be feasible to obtain the needed information by taking a complete enumeration or census of the aggregate. Administrators accustomed to dealing with the censuses were at first inclined to be suspicious of samples and reluctant to use them in place of a census. Although this attitude no longer persists, it may be well to list the principal advantages of sampling as compared with the complete enumeration.

To reduce Cost

If data is secured from only a small fraction of the aggregate, expenditure are smaller than when a complete enumeration is attempted.

With large populations, results accurate enough to be useful can be obtained from samples that represent only a small fraction of the population.

Greater Speed

For the same reason the data can be collected and summarized more quickly with a sample than with a complete enumeration. This is a vital consideration when the information urgently needed.

Greater Scope

In certain types of inquiry highly trained personnel who are limited in their availability, must be used to obtain the data. A complete enumeration is often impracticable: the choice lies between obtaining the information by sampling or not at all. So sampling provides more scope and flexibility regarding the types of information that can be obtained. On the other hand, if accurate information is wanted for many subdivisions of the population, the size of the sample needed to do the job is sometimes so large that a complete enumeration offers the best solution.

Greater accuracy

Because personnel of higher quality can be employed and given intensive training, and because more careful supervision of the fieldwork and processing of results become feasible when the volume of work is reduced, a sample may produce more accurate results than the kind of complete enumeration that can be taken.

II. Cambodia Inetr Censal Population Survey

The Cambodia Inter-Censal Population Survey (CIPS) 2004 is to be conducted in a nationwide representative sample of 21,000 households as the Third Sampling Units (TSUs), 700 Enumeration Areas as the Secondary Sampling Units (SSUs) and 700 villages as the Primary Sampling Units (PSUs).

Several population characteristics were targeted by the study

It was designed to provide information on the following indicators;

Housing and Household amenities

Population and Household characteristics

Sex, age and marital status
Migration status
Literacy / Education level
Economic characteristic
Other relevant information

III. Sampling frame

The General Population Census 1998 databases of the National Institute of Statistics together with the new updated list of villages that were excluded in the general population census of 1998 will be used as the sampling frame for the sampling design of the CIPS 2004.

The frame has the following identification particulars:

1- Province code
2- Province name
3- District code
4- District name
5- Commune code
6- Commune name
7- Village Code
8- Village name
9- Size of village (No. of households)
10- Area code (1 = Urban, 2 = Rural)

IV. Characteristics of the Survey

The sampling design for the CIPS 2004 is a three-stage stratified cluster sampling design, it is a probability sample selection of 100 percent of the Cambodian villages coverage areas, the survey is cover only regular households and will exclude special settlements and institutional households.

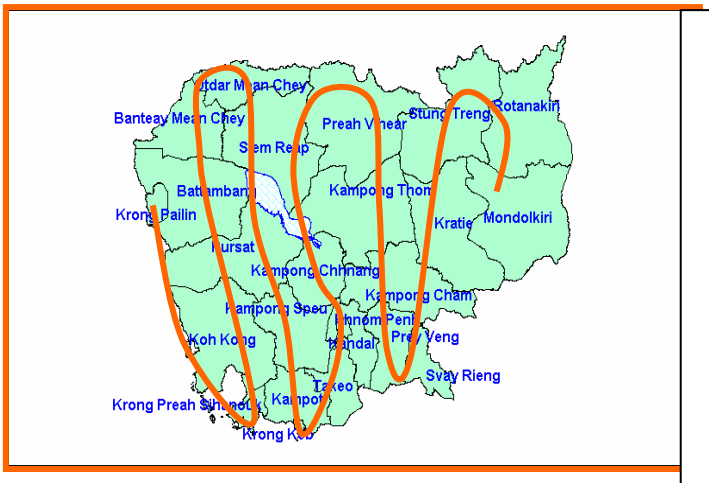
Table-1 : Distribution of number of regular households by province , urban and rural for CIPS 2004

PRO_CODE	PROVINCE_NAME	TOTAL REGULAR HOUSEHOLDS		
		Urban	Rural	Total
01	Banteay Mean Chey	18,201	103,257	121,458
02	Bat Dambang	25,248	145,142	170,390
03	Kampong Cham	7,940	307,097	315,037
04	Kampong Chhnang	7,283	74,425	81,708
05	Kampong Spueu	7,370	110,082	117,452
06	Kampong Thum	12,169	94,050	106,219
07	Kampot	6,007	99,803	105,810
08	Kandal	10,111	192,943	203,054
09	Kaoh Kong	5,292	21,428	26,720
10	Kracheh	14,544	34,142	48,686

11	Mondol Kiri	1,243	4,560	5,803
12	Phnom Penh	94,028	74,387	168,415
13	Preah Vihear	4,102	16,604	20,706
14	Prey Veaeng	11,163	182,141	193,304
15	Pousat	10,704	57,978	68,682
16	Rotanak Kiri	3,124	13,496	16,620
17	Siem Reab	20,131	107,067	127,198
18	Krong Preah Sihanouk	29,314		29,314
19	Stueng Traeng	4,240	9,736	13,976
20	Svay Rieng	4,065	93,714	97,779
21	Takaev	6,900	147,217	154,117
22	Otdar Mean Chey	8,503	14,218	22,721
23	Krong Kaeb	5,236		5,236
24	Krong Pailin	7,510		7,510
TOTAL		324,428	1,903,487	2,227,915

The stratification

The stratification As the urban and rural character of population is to be expected quite different that the sample reflect urban and rural areas proportionately in precise terms and so the sample frame is stratified on this basis prior to selection. Representative it further improved through geographic sequencing of provinces and districts within provinces, again prior to sample selection, this ordering process together with systematic sampling achieves implicit stratification as below.



Snake stile stratification by provinces

ORDER	CODE	PROVINCE NAME
1	11	Mondol Kiri
2	16	Rotanak Kiri
3	19	Stueng Traeng
4	10	Kracheh
5	03	Kampong Cham
6	20	Svay Rieng
7	14	Prey Veaeng
8	06	Kampong Thum
9	13	Preah Vihear
10	17	Siem Reab
11	04	Kampong Chhnang
12	08	Kandal
13	12	Phnom Penh
14	21	Takaev
15	07	Kampot
16	23	Krong Kaeb
17	05	Kampong Spueu
18	15	Pousat
19	22	Otdar Mean Chey
20	01	Banteay Mean Chey
21	02	Bat Dambang
22	18	Krong Preah Sihanouk
23	09	Kaoh Kong
24	24	Krong Pailin

1-The first stage sampling selection

The 700 villages will be selected from the list of village (Frame), which General Population Census 1998 databases conducted by National Institute of Statistics with the new updated list of villages that were excluded in the general population census of 1998 (Updated in 2004 by NIS for Cambodia Inter-Censal Population Survey 2004) the stratification as mention above for provinces and district ready prepared before the sampling selection adopted.
The domain is listed in order of province code, province name, district code, commune code, communes name, village code, village name and the size in the villages known at that time will be its size.

The method of Linear Systematic Sampling with Probability Proportional to Size (LSS-PPS) will be use to select these villages.

The frame would contain in addition to the code for the above identification particulars, the name of the villages as well as the number of households in the village as known at that time and number of households in the village as in the sampling frame, will be its size, to avoid ambiguity, the number of households in a village, as recorded in the frame, will always be referred to as its size.

Example;

The total number of villages in country will be denoted by N and the size of the i-th village in the domain by Si , for i= 1, 2 ,.....,N, and the total cumulative =2227915 households

The number of villages to be selected from the domain by the (LSS-PPS) selection is prepared as a tabular layout using 17 columns and N rows - one for each in the province;

1- Province code
2- Province name
3- District code
4- District name
5- Commune code
6- Commune name
7- Village Code
8- Village name
9- Size of village
10- Area code
11- Lower cumulative (Lo) : $Lo=1+S_1+S_2+.....+S_{(N-1)}$
12- Upper cumulative (Up): $Up=S_1+S_2+……..+S_N$
13- Order of selection
14- The selection interval (I) where I = (UN)/n (UN =2 227 915 total households in the country, n = 700 sample size in the in the survey (Villages))
15- Random start calculation (R) in the range 1 to I by using formula (Rand ()*I)
16- A fix value of random start
17- Output of LSS-PPS calculation R1=R, and generate a sequence of n village
($R_1 = R$, $R_2 = R_1+ I$, $R_3= R_1+ 2I$,……., $R_n= R_1+(n-1)I$)

Table-2 : Distribution of Sample Size of SSUs for CIPS 2004

PRO-CODE	PROVINCES	Sample PSUs (Villages)			
		Combined	Urban	Rural	Total
01	Banteay Mean Chey		5	32	37
02	Bat Dambang	1	8	46	54
03	Kampong Cham	2	2	97	99
04	Kampong Chhnang		3	23	26
05	Kampong Spueu	2	2	34	36
06	Kampong Thum		4	30	34
07	Kampot		1	32	33
08	Kandal	1	3	61	64
09	Kaoh Kong		1	6	7
10	Kracheh		5	11	16
11	Mondol Kiri		1	1	2
12	Phnom Penh		29	23	52
13	Preah Vihear		1	5	6
14	Prey Veaeng		4	57	61
15	Pousat		4	18	22
16	Rotanak Kiri		1	4	5
17	Siem Reab		6	34	40
18	Krong Preah Sihanouk		10	0	10
19	Stueng Traeng	1	1	3	4
20	Svay Rieng		1	29	30
21	Takaev	1	3	46	49
22	Otdar Mean Chey		2	6	8
23	Krong Kaeb		2	0	2
24	Krong Pailin		3	0	3
TOTAL		8	102	598	700

The Sample of Villages (PSUs) would be the yield the sample households that are expected to be interviewed.

The selection probability for village i^{th} will be computed as:

$$P_1 = \frac{n}{A} \times A_i \quad \text{.....(Form . 1)}$$

$$= \frac{700}{2\,227\,915} \times A_i$$

Where: P_1 = Probability of selecting the i^{th} PSU
n = Number of sample villages to be drawn for CIPS
A_i = Number of households in village i^{th} as recorded in the sample frame
A = Total number of households in country as recorded in the sampling frame

2- Second -Stage is a Census Enumeration Area Selection

According to the Cambodia General Population Census 1998 every Sample PSU (Village) was divided into one or more enumeration areas (EAs) according to the size of households in the village with around 110 households per enumeration areas (EA).

The enumeration areas (EAs) were created during cartography conducted in the Cambodia General Population Census 1998, and maps showing the EA boundaries within the villages are available for all the enumeration areas (EAs) with households marked at the National Institute of Statistics (NIS) with the exception of only a few villages.

In this stage, enumeration area (EA) is selected with Simple Random Sampling method.

Table-3 : Distribution of Sample Size of SSUs for CIPS 2004

PRO-CODE	PROVINCES	Sample SSUs (Enumeration Eas)		
		Urban	Rural	Total
01	Banteay Mean Chey	5	32	37
02	Bat Dambang	8	46	54
03	Kampong Cham	2	97	99
04	Kampong Chhnang	3	23	26
05	Kampong Spueu	2	34	36
06	Kampong Thum	4	30	34
07	Kampot	1	32	33
08	Kandal	3	61	64
09	Kaoh Kong	1	6	7
10	Kracheh	5	11	16
11	Mondol Kiri	1	1	2
12	Phnom Penh	29	23	52
13	Preah Vihear	1	5	6
14	Prey Veaeng	4	57	61
15	Pousat	4	18	22
16	Rotanak Kiri	1	4	5
17	Siem Reab	6	34	40
18	Krong Preah Sihanouk	10	0	10
19	Stueng Traeng	1	3	4
20	Svay Rieng	1	29	30

21	Takaev	3	46	49
22	Otdar Mean Chey	2	6	8
23	Krong Kaeb	2	0	2
24	Krong Pailin	3	0	3
TOTAL		102	598	700

The selection probability selection of EA for village i^{th} will be computed as:

$$P_2 = \frac{1}{E_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \text{(Form. 2)}$$

Where: E_i : Total number of enumeration areas (EAs) in the village i^{th}

3- Third-Stage Sampling selection

For the third stage of selection, every selected enumeration area (EA) will be re-listed and mapped in order to provide a current actual number of households frame in the selected enumeration area (EA) for sample selection of 30 households to be interviewed
A fixed sample size of exactly 30 households will be selected by using the method of Linear Systematic Sampling with a random start (LSS).
Where: A random start in range between 01 to Interval (R = 01 to I)

$$I = \frac{M_i^*}{m_i}$$

M^*_i Actual number of households to be listed in a selected EA for village i^{th} at the time of survey
m_i (30 households) the sample households select from the listed of selected EA for village i^{th} at the time of survey
Assume R =R_1 Random start or the 1 st sample household (R = 01 to I)
The 30 sample households are calculate as below

$R = R_1 , R_2 = R_1 + I , R_3 = R_1 + 2I , \ldots\ldots\ldots , R_{30} = R_1 + 29I$

The probability of selecting the households in EA village i^{th}

$$P_3 = \frac{m_i}{M^*_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \text{(Form. 3)}$$

Where:

P_3 = Probability of selecting the households in EA village i^{th}
m_i = (30 households) the sample households select from the listed selected EA for village at the time of survey
M^*_i Actual number of households to be listed in a selected

EA for village i^{th} at the time of survey

Table-4 : Distribution of Sample Size of TSUs for CIPS 2004

PRO-CODE	PROVINCES	Sample TSUs (Households)		
		Urban	Rural	Total
01	Banteay Mean Chey	150	960	1110
02	Bat Dambang	240	1380	1620
03	Kampong Cham	60	2910	2970
04	Kampong Chhnang	90	690	780
05	Kampong Spueu	60	1020	1080
06	Kampong Thum	120	900	1020
07	Kampot	30	960	990
08	Kandal	90	1830	1920
09	Kaoh Kong	30	180	210
10	Kracheh	150	330	480
11	Mondol Kiri	30	30	60
12	Phnom Penh	870	690	1560
13	Preah Vihear	30	150	180
14	Prey Veaeng	120	1710	1830
15	Pousat	120	540	660
16	Rotanak Kiri	30	120	150
17	Siem Reab	180	1020	1200
18	Krong Preah Sihanouk	300	0	300
19	Stueng Traeng	30	90	120
20	Svay Rieng	30	870	900
21	Takaev	90	1380	1470
22	Otdar Mean Chey	60	180	240
23	Krong Kaeb	60	0	60
24	Krong Pailin	90	0	90
TOTAL		3060	17940	21000

4. Substitution of Communes

Some villages of the nation are impossible to survey due to internal insecurity or transportation problems.

It is anticipated, however, that some of the selected villages may also turn out to be difficult to survey, since the security and transportation situation is not static. Should the situation demand it, it will of course be important to accommodate this need in order to protect survey personnel against under danger in the field. Unfortunately, there is no purely scientific way to substitute for sample ex post facto to replace sample units, especially entire PSUs, after they have already been randomly chosen violates probability property of a sample design.

The impact of substitution may, however, be lessened in terms of statistical bias by using as much as practicable, an unbiased substitution procedure design in advance. It is suggested that one way to do this is to choose substitute PSUs, in case where it is mandatory, by designing PSUs

corresponding to one-half the sampling interval, following the original selected PSUs on the geographically stratified computer-sampling frame. It is of course hoped that such substitutions can be kept to a parsimonious minimum.

V. Estimation Procedure

The estimation procedure of CIPS 2004 take account of different weights by PSUs. Thus, the estimation procedure is basically a simple unbiased inflation method, but with a minor adjustment to the number of Enumeration Areas (EAs) comparing to the number of households reliable reported by the village chief, these make sure the number of households listed by the enumerators not under listed.

1. Form A Weight

The probability selection and hence the weight, for the four variables (Total households, Total persons, Total males, Total females) is therefore different than the probabilities and corresponding weights for Form B and the four variables for Form A selected as the two stage sampling selection of first and second stage probabilities :
The Probability of selection (PSUs) as
The Form A sampling weights (design weights) for a village where a full census enumeration area was selected are:

$$w_{1i} = \frac{A}{n \cdot A_i} \cdot \frac{E_i}{1} \quad \text{.....(1)}$$

where:
A = number of households in the country according to census
A_i = number of households in village i according to census
n = number of villages selected
E_i = number of enumeration areas in village i

In the cases where a segment of the selected EA was selected the design weights become:

$$w_{1i} = \frac{A}{n \cdot A_i} \cdot \frac{E_i}{1} \cdot \frac{S_i}{1} \quad \text{.....(2)}$$

S_i = number of segments formed in the selected enumeration area in village i

There is, however, no information from the field regarding the number of segments in the Eas that have been split. In the absence of such information the weights could be calculated as:

$$w_{1i}^{*} = \frac{A}{n \cdot A_i} \cdot \frac{M_i}{M_i^{*}} \quad \text{.....(3)}$$

where:

M_i^* = number of non-vacant households in the selected segment in village *i* according to household listing

M_i = number of households in village *i* according to the village chairman.

Initially, weights were calculated according to (3) for 103 villages. Most of these cases were clear cases of spitted Eas but there were also some cases where "under-listing" of households in the EA was suspected.

The weights according to (3) can be rewritten as:

$$w_{1i}^* = \frac{A}{n \cdot A_i} \cdot \frac{E_i}{1} \cdot \frac{S_i}{1} \cdot \frac{M_i}{E_i \cdot S_i \cdot M_i^*} = w_{1i} \cdot \frac{M_i}{E_i \cdot S_i \cdot M_i^*} \quad \text{.....(4)}$$

Expressed in this way the weights w_{1i}^* are equal to the design weights (2) adjusted with the factor:

$$ADJ_i = \frac{M_i}{E_i \cdot S_i \cdot M_i^*}$$

The factor adjusts the design weight (2) so that the estimate of the total number of households in the village based on form A will agree with the number given by the village chairman.

The adjustments to the design weights can also be done for the the cases where no splitting of EAs have been done. In those cases we have:

$$w_{1i}^* = \frac{A}{n \cdot A_i} \cdot \frac{E_i}{1} \cdot \frac{M_i}{E_i \cdot M_i^*} = w_{1i} \cdot \frac{M_i}{E_i \cdot M_i^*} \quad \text{.....(5)}$$

where:

$$ADJ_i = \frac{M_i}{E_i \cdot M_i^*}$$

An advantage with the adjusted weights (5) as compared to the design weights (1) is that the adjusted weights will show less random variation than the design weights for Form B (see below). It also adjusts for any under-listing of households in the EA or segment.

After discussions within the survey management group it was decided to adjust the design weights using the adjustment factors above in all the 700 villages in the sample. The weights for all villages thus became:

$$w_{1i}^{*} = \frac{A}{n \cdot A_i} \cdot \frac{M_i}{M_i^{*}} \quad \text{.....(6)}$$

2. Sampling Weights for Form B

The sampling weight for household *j* in the sample village *i* will be:

$$w_{ij} = w_{1i}^{*} \cdot w_{2ij} \quad \text{.....(7)}$$

where:

$$w_{2ij} = \frac{M_i^{*}}{m_i}$$

m_i = number of households in the sample from village *i* (m_i =30 for all *i*)

An example

Village 030120808:

Number of households in the country in the census:	2,227,915
Number of villages in the sample:	700
Number of households in the village in the census:	260
Current number of households in the village according to the village chairman:	283
Number of non-vacant households in the selected EA (or segment) listed by the interviewer:	114

The sampling weight for the households in this village becomes:

$$\frac{2{,}227{,}915}{700 \cdot 260} \cdot \frac{283}{114} \cdot \frac{114}{30} = \frac{2{,}227{,}915}{700 \cdot 260} \cdot \frac{283}{30} = 18.7386$$

3. Person weights for Form B

So far we have only discussed household weights. In files where the information is on individual level we need weights for each individual. The household weight could be used as

individual weight (because there has been no random sampling of individuals from the household). All individuals in the household thus get the same weight. It is, however, possible to improve the individual weights by using information from Form A. This has been done in the following way:

a) Estimates of the total number of males and females per stratum (urban/rural * province) were compiled using the total number of males and females in each sampled EA/segment from Form A.

$$\hat{X}^{A} = \sum_{i} w_{1i}^{*} \cdot x_{i}^{A}$$

where:
x_i^A = total number of males/ females in sampled EA/segment *i* according to Form A.

b) Another set of estimates of the total number of males and females per stratum were compiled using the total number of males and females in each sampled household (Form B).

$$\hat{X}^{B} = \sum_{i} \sum_{j}^{30} w_{ij}^{*} \cdot x_{ij}^{B}$$

where:
x_i^B = total number of males/ females in sampled household *ij* according to Form B.

c) Adjustment factors were compiled as:

$$c = \frac{\hat{X}^{A}}{\hat{X}^{B}}$$

The compilation was done for males and females separately in each stratum. This resulted in 2 x 2 x 24 = 96 adjustment factors.

d) The person weights were then compiled as the household weight multiplied by the adjustment factor.

5. Total Population Estimate

As mentioned the Cambodia Inter- Censal Population Survey covered the entire country. Computing the product of basic weight to each individual made an estimate total population of Cambodia variable (total households, total persons, total males, total females) and times the correction factors for each province and take aggregation for the whole country.

VI. Calculations of Sampling Errors

Calculations of sampling errors have been made for some estimates of totals, means and proportions for variables in Form B (annex 3). The software used for the calculations is STATA 8.0. For the calculations presented here we have assumed that stratification was done on provinces and urban/rural (an implicit stratification on province and urban/rural was used for the sample

selection). In seven of the 45 strata there are only one PSU (EA) selected. This causes a problem for the standard error calculations; it is not possible get standard errors in these strata. In these strata we have split the sole EA in two parts and defined the parts as two PSUs.

The standard errors are generally rather small for estimates for major domains like urban/rural and men/women. The coefficients of variation (CV)[1] are below 1 % in many cases. The coefficients of variation are substantially higher for

Provincial estimates, especially for provinces with a small sample (e.g. province 19).

Design effects (Deff)[2] have been calculated for some estimates. They are, as expected, quite low for estimates of demographic characteristics. They are considerably higher for estimates of socio-economic characteristics like employment status (also as expected). For the demographic characteristics "age at first marriage" and "marital status" we find design effects below 5 for major domains like men/women and urban/rural. The socio-economic characteristics are typically more "clustered" than the demographic characteristics, this shows up in generally higher design effects. For the major domain estimates we find design effects up to 20 and occasionally very high values of 200 or more. These "freak" values occur when the sample in terms of number of PSUs is small and when the PSU averages (or proportions) show large variation. One example is the design effect of 285 for the estimate of proportion of government employees in urban areas. The proportion is varying substantially between the 102 PSUs in the domain, the range is from 0 % to 75%.

[1] CV =100* standard error/estimate : $CV = \frac{100 \cdot s.e.(\hat{y})}{\hat{y}}$

[2] Design Effect $Deff = \frac{Variance(\hat{y})}{Variance(\hat{y}srs)}$

2 $CD = \frac{100 \cdot s.e.(\hat{y})}{\hat{y}}$

Sampling Weights and Sampling Errors for the Cambodia Inter-censal Population Survey

Report from a short-term mission to the National Institute of Statistics(NIS),
Phnom Penh, Cambodia

August 19 –26, 2004

by

Hans Pettersson

1. Introduction

The National Institute of Statistics (NIS) has conducted an intercensal demographic survey – *Cambodia Inter-Censal Population Survey 2004 (CIPS)* - that has now been completed in the field. Data entry was completed in May and the data files have been checked and edited.

The purpose of the mission was to:

- Review the CIPS sampling weights using both forms (Form A and Form B).
- Compute CIPS sampling errors.

The Terms of Reference are attached as Annex 1

The mission was conducted in a rather short time (five days). The first task - reviewing the sampling weights - was completed entirely, though it required more time than anticipated. Consequently, there was not sufficient time for computing sampling errors during the mission. We made some test runs for calculation of sampling errors and calculated sampling errors for a few estimates.

2. Review of Sampling weights

The sample design

A three-stage sample design has been used for the CIPS. In the first stage a sample of villages is selected. The villages are implicitly stratified into 45 strata (province*urban/rural). The villages are selected using systematic sampling with probabilities proportionate to size (PPS). The size measures used for the selection was number of households in the village according the the 1998 Census.

In the second stage one Census Enumeration Area was selected randomly (in the head office). At the beginning of the field work all households in the EA were listed. A systematic sample of 30 non-vacant households was selected as the third stage of selection..

The listing of households in the EA will become cumbersome if there are many households in the EA. This may be the case when the enumeration area has grown substantially since the Census. When the EA was large (populationwise) the interviewer was instructed to split the EA into two or more approximately equal-sized segments and to select one segment randomly. All households in the selected segment were listed.

Sampling Weights for Form A

The Form A sampling weights (design weights) for a village where a full census enumeration area was selected are:

$$w_{1i} = \frac{A}{n \cdot A_i} \cdot \frac{E_i}{1} \quad \text{.....(1)}$$

where:
A = number of households in the country according to census
A_i = number of households in village i according to census
n = number of villages selected
E_i = number of enumeration areas in village i

In the cases where a segment of the selected EA was selected the design weights become:

$$w_{1i} = \frac{A}{n \cdot A_i} \cdot \frac{E_i}{1} \cdot \frac{S_i}{1} \quad \text{.....(2)}$$

S_i = number of segments formed in the selected enumeration area in village i

There is, however, no information from the field regarding the number of segments in the Eas that have been split. In the absence of such information the weights could be calculated as:

$$w_{1i}^{*} = \frac{A}{n \cdot A_i} \cdot \frac{M_i}{M_i^{*}} \quad \text{.....(3)}$$

where:
M_i^{*} = number of non-vacant households in the selected segment in village i according to household listing
M_i = number of households in village i according to the village chairman.

Initially, weights were calculated according to (3) for 103 villages. Most of these cases were clear cases of splitted Eas but there were also some cases where “under-listing” of households in the EA was suspected.

The weights according to (3) can be rewritten as:

$$w_{1i}^{*} = \frac{A}{n \cdot A_i} \cdot \frac{E_i}{1} \cdot \frac{S_i}{1} \cdot \frac{M_i}{E_i \cdot S_i \cdot M_i^{*}} = w_{1i} \cdot \frac{M_i}{E_i \cdot S_i \cdot M_i^{*}} \quad \text{.....(4)}$$

Expressed in this way the weights w_{1i}^{*} are equal to the design weights (2) adjusted with the factor:

$$ADJ_i = \frac{M_i}{E_i \cdot S_i \cdot M_i^{*}}$$

The factor adjusts the design weight (2) so that the estimate of the total number of households in the village based on form A will agree with the number given by the village chairman.

The adjustments to the design weights can also be done for the the cases where no splitting of EAs have been done. In those cases we have:

$$w_{1i}^{*} = \frac{A}{n \cdot A_i} \cdot \frac{E_i}{1} \cdot \frac{M_i}{E_i \cdot M_i^{*}} = w_{1i} \cdot \frac{M_i}{E_i \cdot M_i^{*}} \quad \text{.....(5)}$$

where:

$$ADJ_i = \frac{M_i}{E_i \cdot M_i^{*}}$$

An advantage with the adjusted weights (5) as compared to the design weights (1) is that the adjusted weights will show less random variation than the design weights for Form B (see below). It also adjusts for any under-listing of households in the EA or segment.

After discussions within the survey management group it was decided to adjust the design weights using the adjustment factors above in all the 700 villages in the sample. The weights for all villages thus became:

$$w_{1i}^{*} = \frac{A}{n \cdot A_i} \cdot \frac{M_i}{M_i^{*}} \quad \text{.....(6)}$$

Sampling Weights for Form B

The sampling weight for household *j* in the sample village *i* will be:

$$w_{ij} = w_{1i}^{*} \cdot w_{2ij} \quad \text{.....(7)}$$

where:

$$w_{2ij} = \frac{M_i^{*}}{m_i}$$

m_i = number of households in the sample from village i (m_i =30 for all i)

An example

Village 030120808:

Number of households in the country in the census:	2,227,915
Number of villages in the sample:	700
Number of households in the village in the census:	260
Current number of households in the village according to the village chairman:	283
Number of non-vacant households in the selected EA (or segment) listed by the interviewer:	114

The sampling weight for the households in this village becomes:

$$\frac{2,227,915}{700 \cdot 260} \cdot \frac{283}{114} \cdot \frac{114}{30} = \frac{2,227,915}{700 \cdot 260} \cdot \frac{283}{30} = 18.7386$$

Person weights for Form B

So far we have only discussed household weights. In files where the information is on individual level we need weights for each individual. The household weight could be used as individual weight (because there has been no random sampling of individuals from the household). All individuals in the household thus get the same weight. It is, however, possible to improve the individual weights by using information from Form A. This has been done in the following way:

1) Estimates of the total number of males and females per stratum (urban/rural * province) were compiled using the total number of males and females in each sampled EA/segment from Form A.

$$\hat{X}^{A} = \sum_{i} w_{1i}^{*} \cdot x_{i}^{A}$$

where: x_i^A = total number of males/ females in sampled EA/segment *i* according to Form A.

2) Another set of estimates of the total number of males and females per stratum were compiled using the total number of males and females in each sampled household (Form B).

$$\hat{X}^{B} = \sum_{i} \sum_{j}^{30} w_{ij}^{*} \cdot x_{ij}^{B}$$

where: x_i^B = total number of males/ females in sampled household *ij* according to Form B.

3) Adjustment factors were compiled as:

$$c = \frac{\hat{X}^{A}}{\hat{X}^{B}}$$

The compilation was done for males and females separately in each stratum. This resulted in 2 x 2 x 24 = 96 adjustment factors.

4) The person weights were then compiled as the household weight multiplied by the adjustment factor.

3. Calculations of Sampling Errors

Calculations of sampling errors have been made for some estimates of totals, means and proportions for variables in Form B (annex 3). The software used for the calculations is STATA 8.0. For the calculations presented here we have assumed that stratification was done on provinces and urban/rural (an implicit stratification on province and urban/rural was used for the sample selection). In seven of the 45 strata there are only one PSU (EA) selected. This causes a problem for the standard error calculations, it is not possible get standard errors in these strata. In these strata we have split the sole EA in two parts and defined the parts as two PSUs.

The standard errors are generally rather small for estimates for major domains like urban/rural and men/women. The coefficients of variation (CV)[1] are below 1 % in many cases. The coefficients of variation are substantially higher for

[1] CV is =100* standard error/estimate, i.e. : $CV = \frac{100 \cdot s.e.(\hat{y})}{\hat{y}}$

provincial estimates, especially for provinces with a small sample (e.g. province 19).

Design effects (Deff) have been calculated for some estimates. They are, as expected, quite low for estimates of demographic characteristics. They are considerably higher for estimates of socio-economiccharacteristics like employment status (also as expected). For the demographic characteristics "age at first marriage" and "marital status" we find design effects below 5 for major domins like men/women and urban/rural. The socio-economic characteristics are typically more "clustered" than the demographic characteristics, this shows up in generally higher design effects. For the major domain estimates we find design effects up to 20 and occasionally very high values of 200 or more. These "freak" values occur when the sample in terms of number of PSUs is small and when the PSU averages (or proportions) show large variation. One example is the design effect of 285 for the estimate of proportion of government employees in urban areas. The proportion is varying substantially between the 102 PSUs in the domain, the range is from 0 % to 75%.

TERMS OF REFERENCE
Request for Mission of Sampling Expert for CMB/01/P04 from the Royal Government of Cambodia

Title: Sampling Expert
Duration: one week
Dates required: June or July 2004
Duty Station: NIS/Ministry of Planning, Phnom Penh

Duties:

Under the guidance of the Director General of the National Institute of Statistics (NIS), Ministry of Planning, Royal Government of Cambodia and with the national and international staff attached to the NIS, the Consultant will be expected to advise and assist the Government and participate in the following activities for the preparation of Cambodia's Inter-Censal Population Survey *(CIPS)* conducted in March 2004:

1. Review the CIPS Sampling Weights using both forms (Form A and Form B).
2. Compute the CIPS Sampling Error.

Expected Outputs:
Upon the completion of the mission, the consultant is required to produce the following outputs and submit to UNFPA:
- Mission report outlining the mission activities, achievement of work done

List of persons met

NIS:
Mr San Sy Than, Director General
Mrs Hang Lina, Deputy Director
Mr Has Bunton, Deputy Director
Mr Sok Kosal, Bureau Chief of Industry and Trade
Mr Yem Suong ,Director of Social Economic Department
Mr They Kheamm, Deputy Director of Census and Survey Department
Mr Meng Kimkor
Mr Vy Bossadine
Mr Meas Rathmany

UNFPA
Mr Bjarke Oxlund
Mr Chap Rathana

Consultants
Dr Nott Rama Rao
Mr Harry Lode

Standard errors, confidence intervals and corfficients of variation for selected estimates.

Estimate: Total population

Pweight:	Personsweight	Number of obs =	102558
Strata:	Strat	Number of strata =	45
PSU:	PSU	Number of PSUs =	700
		Population size =	12824252

Province	Estimate	SE	95% Confidence Interval		CV(%)[2]
			Lower	Upper	
1	678,882	19,018	641,608	716,157	2.8
2	971,894	42,085	889,408	1,054,380	4.3
3	1,655,349	31,248	1,594,102	1,716,596	1.9
4	531,516	55,417	422,900	640,133	10.4
5	676,821	18,190	641,167	712,474	2.7
6	606,918	27,055	553,890	659,945	4.5
7	596,200	13,075	570,572	621,827	2.2
8	1,203,134	30,473	1,143,407	1,262,861	2.5
9	118,495	4,415	109,843	127,148	3.7
10	328,885	20,966	287,792	369,978	6.4
11	37,048	346	36,369	37,726	0.9
12	1,043,669	45,690	954,116	1,133,222	4.4
13	150,220	13,255	124,241	176,199	8.8
14	1,013,087	24,694	964,686	1,061,488	2.4
15	455,793	62,504	333,284	578,301	13.7
16	100,248	12,880	75,003	125,494	12.8
17	755,486	22,854	710,692	800,280	3.0
18	186,762	10,121	166,925	206,599	5.4
19	89,923	2,767	84,500	95,346	3.1
20	513,616	15,346	483,536	543,695	3.0
21	880,405	21,504	838,257	922,553	2.4
22	130,491	10,784	109,354	151,629	8.3
23	58,165	11,167	36,277	80,052	19.2
24	41,248	514	40,240	42,255	1.2

[2] CV(%) is =100* standard error/estimate, i.e. : $CV = \frac{100 \cdot s.e.(\hat{y})}{\hat{y}}$

Estimate: Percent literate persons, by province

	Estimate	SE	95% Confidence Interval		CV(%)
			Lower	Upper	
p==1	62.0	1.7	58.6	65.3	2.8
p==2	62.9	1.2	60.5	65.3	1.9
p==3	62.5	1.7	59.3	65.8	2.6
p==4	55.8	2.1	51.8	59.9	3.7
p==5	58.9	2.9	53.2	64.5	4.9
p==6	60.5	3.2	54.1	66.8	5.3
p==7	63.4	2.2	59.0	67.8	3.5
p==8	69.3	1.2	66.9	71.8	1.8
p==9	56.8	5.5	46.0	67.6	9.7
p==10	55.0	4.9	45.3	64.7	8.9
p==11	58.2	4.3	49.8	66.5	7.3
p==12	80.2	0.9	78.3	82.1	1.2
p==13	54.5	5.5	43.7	65.2	10.0
p==14	63.2	1.4	60.4	66.0	2.3
p==15	58.8	2.4	54.0	63.5	4.1
p==16	20.5	2.7	15.2	25.7	13.2
p==17	54.6	2.9	48.9	60.4	5.4
p==18	67.9	3.5	61.1	74.7	5.1
p==19	49.5	14.4	21.2	77.8	29.2
p==20	69.6	1.2	67.3	72.0	1.7
p==21	65.1	1.3	62.5	67.6	2.0
p==22	57.7	2.3	53.1	62.2	4.0
p==23	58.4	3.9	50.8	66.1	6.7
p==24	64.0	8.4	47.5	80.4	13.1
All	63.2	0.5	62.2	64.2	0.8

Estimate: Percent literate persons, by sex and urban/rural

	Estimate	SE	95% Confidence Interval		CV(%)
			Lower	Upper	
Men	68.8	0.5	67.7	69.8	0.8
Women	58.1	0.6	57.0	59.2	1.0
Urban	72.8	1.0	70.9	74.7	1.4
Rural	61.6	0.6	60.4	62.7	0.9
All	63.2	0.5	62.2	64.2	0.8

Estimate: Average age at first marriage

	Estimate	SE	95% Confidence Interval		CV(%)
			Lower	Upper	
Men	23.3	0.05	23.2	23.5	0.2
Women	20.8	0.05	20.7	20.9	0.2
Urban	22.5	0.10	22.4	22.7	0.4
Rural	21.8	0.05	21.7	21.9	0.2
All	21.9	0.04	21.8	22.0	0.2

Estimate: Marital status, per cent of population

		Estimate	SE	95% Confidence Interval		CV(%)
				Lower	Upper	
Men	Never married	61.5	0.20	61.1	61.9	0.3
	Married	37.0	0.20	36.6	37.4	0.5
	Widowed	0.9	0.05	0.8	1.0	5.3
	Divorced	0.4	0.03	0.4	0.5	7.0
	Separated	0.1	0.02	0.1	0.1	14.5
Women	Never married	53.7	0.20	53.3	54.1	0.4
	Married	37.3	0.21	36.9	37.7	0.6
	Widowed	6.4	0.15	6.1	6.7	2.3
	Divorced	2.0	0.08	1.8	2.1	4.1
	Separated	0.6	0.05	0.5	0.7	8.4

Estimate: Sector of employment, per cent of all economically active

		Estimate	SE	95% Confidence Interval		CV(%)
				Lower	Upper	
Urban	Government	13.0	0.92	11.1	14.8	7.1
	Private	85.5	1.03	83.5	87.5	1.2
Rural	Government	2.9	0.14	2.6	3.2	4.8
	Private	96.4	0.22	95.9	96.8	0.2
All	Government	4.2	0.17	3.9	4.5	4.0
	Private	94.9	0.23	94.5	95.4	0.2

Estimate: Sector of employment, per cent of all economically active

		Estimate	SE	95% Confidence Interval		CV(%)
				Lower	Upper	
Men	Government	7.0	0.28	6.5	7.5	4.0
	Private	91.9	0.35	91.2	92.6	0.4
Women	Government	1.6	0.12	1.4	1.8	7.5
	Private	97.7	0.17	97.4	98.1	0.2
All	Government	4.2	0.17	3.9	4.5	4.0
	Private	94.9	0.23	94.5	95.4	0.2

Estimate: Employment status, per cent of all economically active

		Estimate	SE	95% Confidence Interval		CV(%)
				Lower	Upper	
Men	Paid employee	16.6	0.51	15.6	17.6	3.1
	Own account	49.9	0.52	48.9	51.0	1.0
	Family worker	33.0	0.55	31.9	34.1	1.7
Women	Paid employee	9.6	0.4	8.8	10.4	4.2
	Own account	28.5	0.55	27.4	29.6	1.9
	Family worker	61.5	0.71	60.1	62.9	1.2
Urban	Paid employee	30.7	1.44	27.9	33.6	4.7
	Own account	39.7	0.85	38.0	41.4	2.1
	Family worker	28.6	1.46	25.8	31.5	5.1
Rural	Paid employee	10.3	0.42	9.5	11.1	4.1
	Own account	38.7	0.34	38.1	39.4	0.9
	Family worker	50.6	0.54	49.6	51.7	1.1
All	Paid employee	13.0	0.41	12.2	13.8	3.2
	Own account	38.9	0.32	38.2	39.5	0.8
	Family worker	47.8	0.51	46.8	48.7	1.1

Design effects for selected estimates

Estimate: Average age at first marriage

	Estimate	SE	Deff
Men	23.3	0.05	1.8
Women	20.8	0.05	1.7
Urban	22.5	0.10	1.8
Rural	21.8	0.05	2.5
All	21.9	0.04	2.4

Estimate: Marital status, per cent of population

		Estimate	SE	Deff
Men	Never married	61.5	0.20	1.9
	Married	37.0	0.20	2.7
	Widowed	0.9	0.05	5.5
	Divorced	0.4	0.03	4.5
	Separated	0.1	0.02	4.8
Women	Never married	53.7	0.20	1.9
	Married	37.3	0.21	2.7
	Widowed	6.4	0.15	5.5
	Divorced	2.0	0.08	4.5
	Separated	0.6	0.05	4.8

Estimate: Sector of employment, per cent of all economically active

		Estimate	SE	Deff
Urban	Government	13.0	0.92	285.5
	Private	85.5	1.03	60.4
Rural	Government	2.9	0.14	4.6
	Private	96.4	0.22	1.9
All	Government	4.2	0.17	4.1
	Private	94.9	0.23	6.2

Estimate: Sector of employment, per cent of all economically active

		Estimate	SE	Deff
Men	Government	7.0	0.28	13.0
	Private	91.9	0.35	2.7
Women	Government	1.6	0.12	9.6
	Private	97.7	0.17	0.6
All	Government	4.2	0.17	4.1
	Private	94.9	0.23	6.2

Estimate: Employment status, per cent of all economically active

		Estimate	SE	Deff
Men	Paid employee	16.6	0.51	20.1
	Own account	49.9	0.52	8.2
	Family worker	33.0	0.55	12.6
Women	Paid employee	9.6	0.4	18.8
	Own account	28.5	0.55	13.3
	Family worker	61.5	0.71	13.2
All	Paid employee	13.0	0.41	8.5
	Own account	38.9	0.32	2.4
	Family worker	47.8	0.51	5.9