VIET NAM LIVING STANDARDS SURVEY 1997/98

Revised Sample Design

Chris Scott, Consultant. July 1997

1. BACKGROUND

This report results from a 1-week consultancy assignment in Viet Nam from July 7 to 12, 1997. The Consultant worked with the staff of the General Statistics Office, Department of Social-Environmental Statistics, under the guidance of Mr Nguyen Van Tien, and in close collaboration with the project's technical group.

I wish to thank the above officials and their staff for their full and effective cooperation. Their contribution to this report was crucial.

The Consultant had made a similar visit in November 1991 to assist in the preparation of the sample design for the initial Viet Nam Living Standards Survey, carried out in 1993.[1]

When the currently planned survey, the LSS-97, was proposed the Consultant made a 2-week visit to Hanoi in May 1996 to assist in the preparation of a provisional sample design.[2] The sample size proposed at that time was 8,000 households. However, the start of the survey was delayed and in 1997 the Government decided to limit the sample to 6,000 households. This decision required a substantial revision of the initial 1996 sample design. The present report gives the definitive design.

The changes result not only from the reduction in the sample size but also from an error in the initial 1996 report: that report was based on the understanding that the 1993 survey sample of communes had been selected with probabilities proportional to their census population: in reality, the communes had been selected with probabilities proportional to their number of census *households*. This, and some other minor errors, have been corrected in the present up-date.

---

[1] See report "Viet Nam Living Standards Survey: Recommended Sample Design". Chris Scott, November 1991, World Bank.

[2] See report "Viet Nam Living Standards Survey 1996/97: Recommended Sample Design." Chris Scott, May 1996, World Bank.

2. GENERAL DESIGN OF THE SURVEY

The 1993 LSS was based on a nationally representative sample of 4800 households, interviewed from October 1992 to October 1993. The present project is essentially a repeat survey, using a somewhat larger sample, with as many as possible of the households of the earlier survey included again to allow panel comparisons over a 5 year interval.

The new survey is to be spread over 12 months, beginning late 1997, using the LSMS-type design, based on 2 interviews with each household, spaced at a 2-week interval.

The full sample of 6000 households should be planned to allow disaggregation of results into the 7 Viet Nam regions for the rural sector plus 3 urban domains. Table 1 lists these 10 domains with their estimated populations and household numbers in 1994.

TABLE 1. The 10 Survey Domains

DOMAIN	Resident population 1994 ('000)	Households ('000)
URBAN		
1. Ha Noi, Ho Chi Minh	4 403	897
2. Other cities	3 911	797
3. Other urban population	5 826	1 187
TOTAL URBAN	14 140	2 881
RURAL		
4. North Mountain	10 753	2 129
5. Red River Delta	11 578	2 763
6. North Central	8 673	1 822
7. Central Coast	5 796	1 112
8. Central Highlands	2 312	390
9. South East	4 786	867
10. Mekong Delta	13 429	2 496
TOTAL RURAL	57 327	11 579

3. STRUCTURE OF THE SAMPLE

Following the dual objectives of the survey, the sample falls into two parts. First, it is desired to re-interview as many households as possible among those interviewed in the Living Standards Survey of 1993 (LSS93). Second, a representative national sample of 6000 households is required overall. Thus, there is a first part of the sample that comes from LSS93 and a second part which completes the representative 6000. It was decided by the GSO that the second part should be drawn from the current national survey carried out by the same department of the GSO entitled the Multi-Purpose Household Survey (MPHS).

Both of these sources will be described in more detail in Section 4 below. But before doing so we need to define rather more precisely the distinction made above between the two parts of the sample. The problem is that the set of households available for interview today which were interviewed in 1993 is an unrepresentative subsample. This is because some of the households interviewed in 1993 have disappeared: these are likely to be the more mobile, or the more unstable, households. Those that remain (which may be called the "Panel") are therefore not fully representative. In order to make up a representative sample we need to replace these missing households. Thus in reality we have <u>three</u> parts to the sample, not two:

(1) The Panel (households interviewed in 1992/93 and available for interview in 1997/8).

(2) The replacements needed to correct the drop-out from the Panel.

(3) The subsample, nationally representative, from the MPHS needed to complete the sample of 6000.

Where should we get subsample (2)? It would be possible to obtain these from the MPHS also, but it is much simpler to get them from the same place where they disappeared, namely from the village or precinct in which they were interviewed in 1992/93. We only need to make the rule that the households interviewed in 1997/98 in each village/precinct should be made up to a constant number by replacing those who dropped out between 1992/93 and 1997/98. (As we shall see below, this number will be 16.) With this rule, the samples (1) and (2) together will become representative, and sample (3) will be representative, and we can simply put these three parts together. Sample (2) does not enable us to have a representative panel, but it does enable us to make use of the panel households in the national sample.

4. SUMMARY DESCRIPTIONS OF THE LSS93 AND MPHS SAMPLES

In order to construct the desired sample of 6000 households it is necessary to look carefully at the design of the two source samples, the LSS93 and the MPHS.

4.1 LSS93 Sample Design

The LSS93 sample was selected in 3 stages: communes, villages/precincts, households.

At the 1st stage, 150 communes were selected with probability proportional to census households. The selected communes were asked, by letter, to report the villages/precincts they contained and the number of households in each.

At the 2nd stage, 2 villages (rural) or precincts (urban) were selected in each commune with probability proportional to the number of households reported by the commune authorities.

The survey teams then visited each selected village/precinct and listed all households. From these they selected 16 households with equal probability.

The total sample comes to 150 x 2 x 16 = 4800 households, the sample being roughly self-weighting ("representative" or "EPSEM"). In the analyses carried out on this sample it has always been treated as self-weighting. Essentially, this assumes that the 1989 census was still valid in 1993. This consultant considers this assumptions to be acceptable in the circumstances.

4.2 MPHS Sample Design

The basic structure of the sample is the same as that of the LSS93, described above.

The principal difference was at the 1st stage. In the MPHS, the provinces were first classified by population size into 7 size-strata and different sampling fractions were fixed for allocating the 1st stage sample numbers (communes) between these strata. The total size of the 1st stage sample was 10 times as large (1500 compared with 150 communes in LSS93). The 1st stage units, the communes, were then selected within provinces with probabilities proportional to their census population.

The 2nd stage was similar to that in LSS93 except that this time the "size" of the villages/precincts was measured by their population, as reported by the commune, instead of their number of households.

The 3rd stage differed only in that the number of households selected per village/precinct was 15 instead of 16.

Sample size was 1500 x 2 x 15 = 45 000 households.

5. SAMPLE ALLOCATION BETWEEN THE DOMAINS

The first step is to determine the number of communes to be allocated to each of the 10 sampling domains.

Returning to the sample for the current survey, the procedure will be first to fix the domain allocations for the sample desired (the 6000 households), then to determine the allocation of the LSS93 sample, then to subtract the latter from the former to find how many households need to be selected in each domain from the MPHS.

5.1 Allocation for the Desired Sample

If we sample with equal probability in every domain the sample of 6000 households will be distributed as shown in Table 2. Here we distribute the 6000 households in proportion to the number of census households shown in Table 1. This would be the equal probability or "EPSEM" sample distribution, with equal sampling fractions (weights) in every domain.

Is this a satisfactory distribution, or do we need to introduce weights to ensure a large enough sample in each domain for the needs of the analysis?

A rough, but reasonable, working rule is that we need a sample of approximately 500 households in any domain to allow for an adequate analysis. We see that 6 of the 10 domains fail to meet this requirement. Yet the total of 6000 households would be theoretically enough to allow 500 per domain. What can be done to achieve a more acceptable distribution of the sample?

All that is needed is to vary the sampling fractions among the domains. The resulting bias in the sample can be put right by re-weighting. For example, if one domain has twice as high a sampling fraction we can give it a weight of one-half in the analysis. It pays to use simple weights, to use as few different weights as possible, and to avoid extreme differences between the largest and smallest weights. After experimentation with a variety of weighting schemes, the set of domain sampling fractions (= 1/W) shown in Table 3 was chosen by the Project Director as one that meets most of these objectives.

TABLE 2. Sample Allocation (I): Proportional

DOMAIN	SAMPLE SIZE (Households)
URBAN	
1. Ha Noi, Ho Chi Minh	372
2. Other cities	331
3. Other urban population	493
TOTAL URBAN	1196
RURAL	
4. North Mountain	883
5. Red River Delta	1146
6. North Central	756
7. Central Coast	461
8. Central Highlands	162
9. South East	360
10. Mekong Delta	1036
TOTAL RURAL	3904
NATIONAL TOTAL	6000

TABLE 3. Sample Allocation (II): Preferred relative sampling fractions per domain

DOMAIN h	RELATIVE SAMPLING FRACTION $F_h=K/W_h$ (see note)	SAMPLE SIZE (Households)
URBAN		
1. Ha Noi, Ho Chi Minh	2	568
2. Other cities	2	505
3. Other urban population	1.5	564
TOTAL URBAN		1637
RURAL		
4. North Mountain	1	674
5. Red River Delta	1	874
6. North Central	1	577
7. Central Coast	1.5	528
8. Central Highlands	3	371
9. South East	2	549
10. Mekong Delta	1	790
TOTAL RURAL		4363
NATIONAL TOTAL		6000

NOTE. Sampling fraction F_h is the ratio sample size to population size in domain h. Weight $W_h = 1/F_h$ is the weight used in the analysis to correct for unequal sampling fractions in the domains. Constant K is inserted to ensure overall sample size fixed (6000).

In estimates relating to groups of domains, or to the whole sample, the data will be weighted by the factors shown in order to correct the bias due to deliberate over- or under-sampling. This manipulation increases the sampling error for such estimates but

this is the inevitable price to pay for ensuring an adequate sample for all domains. Thus, the advantage regarding the smaller domains is paid for by a fairly small disadvantage at the wider level.

The above statement regarding corrective weighting is, however, somewhat theoretical in the present case. It would apply to a sample determined solely by the above weights. In the present case the sample is put together from two sources with different sample designs and the weights shown are only averages. The correct weights to apply are shown in Section 7 of this report.

5.2 Distribution of the LSS93 sample

Next we have to look at the allocation of the LLS93 sample among the domains. Referring back to page 2 above, what is needed here is the sum of the two samples (1) and (2). This is obtained from the number of communes in LSS93 multiplied by 32. The distribution is shown in the first column of Table 4.

5.3 Distribution of the MPHS subsample

The next step is to subtract the sample numbers shown in the left-hand column of Table 4 from the "target" figures shown in the right-hand column of Table 3. The difference is the number of households which are required to be selected from the MPHS sample. This is shown in Table 4. These numbers have to be complete communes, so that some rounding is necessary. It will be recalled that 30 households are selected from each commune in the MPHS. Thus Table 4 shows in the right-hand column the rounded number of households followed by the number of communes to be selected in each domain from the MPHS.

It will be seen that in one domain (5 - Red River Delta) the LSS93 sample exceeds the target by 86 households. In this instance, instead of seeking a supplementary sample we have to eliminate a small part of the sample from LSS93 - to be exact, three communes need to be eliminated from the LSS93 sample in this domain. It is proposed to deal with this anomaly by re-weighting the data from the remaining 27 communes in the LSS93 sample in this domain. The required weight will be 30/27.

TABLE 4. Sample Allocation in the LSS93 sample and supplementary sample needed from MPHS

DOMAIN	LSS93 h'holds	Target total h'holds	Supplement Needed h'holds	Supplement Adjusted	Supplement Communes
URBAN					
1. Ha Noi, Ho Chi Minh	288	568	280	270	9
2. Other cities	288	505	217	210	7
3. Other urban	416	564	148	150	5
TOTAL URBAN	992	1637	645	630	21
RURAL					
4. North Mountain	672	674	2	0	
5. Red River Delta	960-96	874	-86	0	
6. North Central	576	577	1	0	
7. Central Coast	384	528	144	150	5
8. Central Highlands	128	371	243	240	8
9. South East	320	549	229	240	8
10.Mekong Delta	768	790	22	30	1
TOTAL RURAL	3808-96	4363	555	660	22
NATIONAL TOTAL	4800-96	6000	1200	1290	43

While Table 4 lays down the number of communes and households required in the subsample from the MPHS, this is not the end of the

story. It will be recalled that the MPHS sample was selected with probabilities that varied from one province to another within the same domain. Such varying sampling rates between provinces within one domain are not required for the MPHS subsample. It is therefore still necessary to work out the sample allocation between provinces for this subsample. This allocation, within each domain, has to be proportional to the province populations within the domains.

The necessary data have been set out in Table 5.[3] The right-hand section of the table shows the allocation of the sample communes among the provinces. This table requires some commentary.

The stub on the left sets out the different domains and provinces, distinguishing the three urban domains and the rural. Cols.1, 2, 3 and 4 give the 1995 estimated population figures for each of these units.

Cols. 5, 6 and 7 show the number of communes selected into the MPHS total sample.

Cols. 8, 9 and 10 repeat these data after eliminating those communes which coincide with a commune of the LSS93 sample. The MPHS subsample has to be selected from among the communes in these columns.

Cols. 11, 12 and 13 show the allocation of the subsample of communes among the units listed. This allocation is worked out using the method normally used for systematic sampling with probability proportional to size, thus ensuring a good spread of the subsample communes across the provinces of each rural domain (or cities and towns of each urban domain). The method is to cumulate the population across these units in the domain, to compute the sampling (i.e. allocation) interval for the domain, using the numbers of communes taken from Table 4, to select a random number not greater than the interval, and to sample at the stated interval in the cumulative column. All this is done separately for each domain. The table shows the cumulations and the resulting allocations. The supporting Table 5A shows the sample sizes (number of communes), intervals and random starts for each domain. In order to profit from this form of allocation, the provinces or towns within each domain were arranged, prior to selection, in geographical order (North-South, or serpentine order where appropriate): this ensures that the achieved spreading of the sample across the list of units corresponds to a spreading of the sample across the national terrain. The ordering of the units within domains in Table 5 is in fact the order used in sample allocation.

---

[3] The extensive work for this table was carried out by the project staff at GSO, Dept. of Social-Environmental Statistics.

TABLE: 5 ALLOCATION OF MPHS SUBSAMPLE COMMUNES BY DOMAIN AND PROVINCE

No	Region, province (Town)	Population (1994)				MPHS communes			Less overlap			Cummulative Pop.			Sample		
		Majo cities	Other cities	Other Urban	Rural	Cities	Other Urban	Rural	Cities	Other Urban	Rural	Cities	Other Urban	Rural	Cities	Other Urban	Rural
A	*B*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
	North Mountain	**0**	**453**	**1183**	**10752**	**25**	**48**	**267**	**24**	**47**	**263**				**1**	**1**	**0**
1	Ha Giang			48	488		4	20		4	19		48	488			
2	Cao Bang			65	573		4	20		4	20		113	1061			
3	Lai Chau			74	446		5	19		5	19		187	1507			
4	Loa Cai			73	479		5	19		5	19		260	1986			
5	Tuyen Quang			64	581		4	20		4	20		324	2567			
6	Lang Son			90	600		5	19		5	19		414	3167			
7	Bac Thai(Thai Nguyen)		186	34	948	8		20	7		20	186	448	4115	1		
8	Yen Bai			116	536		6	18		6	18		564	4651		1	
9	Son La			103	699		5	19		4	19		667	5350			
10	Vinh Phu(Viet Tri)		125	38	2085	5		31	5		30	311	705	7435			
11	Quang Ninh(Ha Long)		142	245	512	12		12	12		11	453	950	7947			
12	Ha Bac			115	2193		4	32		4	32		1065	10140			
13	Hoa Binh			118	612		6	18		6	17		1183	10752			
	Red River Delta	**1151**	**784**	**554**	**11580**	**39**	**17**	**181**	**39**	**17**	**175**				**3**	**1**	**0**
14	Ha Noi City	1151				19			19			1511			2		
15	Ha Noi Other			0	1044			17			16			1044			
16	Hai Phong		540	0	1076	13		19	13		18	993		2120	1		
17	Ha Tay			173	2084		5	31		5	30		1356	4204			
18	Hai Hung			141	2568		4	35		4	35		1497	6772			
19	Nam Ha(Nam Dinh)		244	59	2337	7		31	7		31	1237	1556	9109			
20	Thai Binh			104	1686		4	28		4	26		1660	10795			
21	Ninh Binh			77	785		4	20		4	19		1737	11580		1	

TABLE 5. (Cont.)

A	B	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	South East	**3252**	**519**	**321**	**4786**	**46**	**7**	**97**	**46**	**7**	**95**				**8**	**0**	**8**
39	Ho Chi Minh city	3252				25			25			3252			7		
40	Ho Chi Minh Other				1140			17			17			1140			2
41	Song Be			55	1059		3	25		3	25		3627	2199			1
42	Tay Ninh			105	783		4	20		4	19		3732	2982			2
43	Dong Nai(Bien Hoa)		360	95	1358	11		21	11		21	3267	3827	4340	1		2
44	Ba Ria-VT(Vung Tau)		159	66	446	10		14	10		13	3426	3893	4786			1
	Mekong River Delta	**0**	**485**	**1938**	**13430**	**15**	**58**	**247**	**15**	**57**	**241**				**1**	**2**	**1**
45	Long An			158	1094		5	23		5	23		4051	1094			
46	Dong Thap			236	1254		7	21		7	21		4287	2348		1	
47	An Giang			366	1605		9	23		9	23		4653	3953			
48	Tien Giang(My Tho)		165	36	1455	6		26	6		26	3591	4689	5408	1		
49	Kien Giang			285	1074		9	19		9	18		4974	6482			
50	Can Tho		320	33	1464	9		23	9		22	3911	5007	7946			
51	Ben Tre			98	1233		4	24		4	24		5105	9179			
52	Vinh Long			145	917		5	23		5	22		5250	10096		1	
53	Tra Vinh			59	900		3	21		3	20		5309	10996			
54	Soc Trang			196	1002		7	21		6	20		5505	11998			1
55	Minh Hai			326	1432		9	23		9	22		5831	13430			
	TOTAL	**4403**	**3911**	**5831**	**57327**	**186**	**189**	**1125**	**182**	**184**	**1103**				**16**	**5**	**22**

71472

TABLE 5A. Parameters used for commune sample allocation in Table 5

DOMAIN	PARAMETERS No. of communes	Selection interval	Random start
URBAN			
1. Ha Noi	2	-	-
Ho Chi Minh City	7	-	-
2. Other cities	7	559	127
3. Other urban	5	1165	559
TOTAL URBAN	21		
RURAL			
4. North Mountain	0	-	-
5. Red River Delta	0	-	-
6. North Central	0	-	-
7. Central Coast	5	1159	1016
8. Central Highlands	8	289	159
9. South East	8	598	442
10. Mekong Delta	1	-	11253
TOTAL RURAL	22		

00

6. SAMPLE SELECTION

0

6.1 Commune selection

The communes of the LSS93 sample are already selected. The only remaining task is the elimination of the 3 surplus communes in Domain 5 (RED River Delta). This was done by systematic selection with equal probability, selecting 3 from the 30 sample communes, for rejection.

The communes of the MPHS subsample still have to be selected. Table 5 shows how many communes need to be selected in each city or province (see columns 14, 15 and 16). These have to be selected from those enumerated in columns 8, 9 and 10 of Table 5. For example, in the rural part of Domain 7 (Central Coast Region), in the Province of Quang Ngai, 1 rural commune (col.16) has to be selected from among the 22 shown in col.10. These 22 consist of all the MPHS rural communes selected in that province after excluding the overlap with LSS93. Just 1 of the 22 is required to complete the LSS97 sample.

This subsample must be selected with equal probability within any given province or city. Since the MPHS communes have already been selected with probability proportional to size (PPS) in the 1994 MPHS sample selection, the PPS sampling does not need to be repeated. By subsampling with equal probability we obtain a PPS subsample.

The subsampling should be carried out by systematic equal probability sampling. Number the available MPHS communes from 1 to N. If n are to be selected, use the sample interval N/n. Begin with a random number between 1 and N/n: add the interval repeatedly to select the desired sample. If only 1 has to be selected, use a random number between 1 and N to indicate the selected commune.

In the above example, 1 commune has to subsampled from 22. Here N = 22. Select a random number between 1 and 22; suppose this is 12. Then select the 12th commune among the 22 MPHS sample communes identified in column 10.

The selection of communes in the two subsamples, merged into a single sample list, had been completed by the Project staff by the end of this Consultant's visit.

6.2 Village/precinct selection

Both in the LSS93 sample and the MPHS sample, all sample communes have already 2 villages or precincts selected (with PPS). Thus no further work is needed on village/precinct selection.

6.3 Household selection

In the LSS93 sample we have all the panel households, all of which are to appear in the new LSS97 sample. In the MPHS, all selected villages already have 30 households selected, so no action is required on these.

It only remains to select the replacement households, which make up the gap between the panel sample and the full list of 16

households per village/precinct in the LSS sample. This gap is the sample (2) mentioned on p.3 of this report.

The size of this gap can only be known at the time when the survey teams go into the LSS93 communes for the new survey. Before starting to interview, the team has to determine the number of the original households which are still available (either at their original dwelling or elsewhere within the village). They do this by visiting the dwellings of the LSS93 sample households. Subtracting this number from 16 in each village/precinct gives the number still needed, i.e. the replacement sample. However, a number of these will be "automatic replacements", that is, households who have moved into the dwelling vacated by the sample household which has disappeared. Just possibly, a sample household may have divided into two: in this case both have to be interviewed in 1996/7, thus reducing by 1 the number of replacements needed. When the correct number is known, the team selects the necessary replacements by the following procedure.

The team first asks the village authorities for a list of all households in the village. The team supervisor numbers these households from 01 upwards. He then uses his table of random numbers to select a replacement household. (If the number of households in the village is less than 100 he uses a 2-digit random number. If the random number exceeds the number listed, he selects another. If it falls on a household already selected, he selects another. In this way he selects all the replacement households needed.)

It is possible, though very unlikely, that enough households have split into 2 to make a total, with automatic replacements, that exceeds 16 in one village. In this case use the random number table to eliminate the one or two excess households at random until exactly 16 is achieved.

7. COMPUTATION OF WEIGHTS FOR USE IN ESTIMATION

The sample designs for the LSS93 and MPHS have been described above in Section 4.

The LSS93 design is taken over in its entirety, with updating at the household level, into the LSS97, with the exception of the 3 communes of Domain 5 which have to be dropped. The latter have been identified - see Section 6.1 above.

The remainder of the LSS97 is the subsample from the MPHS selected at the commune level (43 communes out of 1500).

The LSS93 was designed to be approximately self-weighting (i.e. no weights required). But time has passed and populations

have changed in 4 years. Corrective weighting is now desirable if the sample is to be extrapolated to the national level.

As for the MPHS, this was not designed as self-weighting. The sample allocation to each province is not in proportion to population or household numbers. The subsampling from the MPHS for the LSS97 is designed to achieve a distribution ensuring an adequate sample in every domain (with the exception of Domain 8 in which a smaller sample has been allowed because of the heavy expense involved in field work in that region). There is also the uncertain assumption that the census population updated at province level in 1993 can be regarded as an adequate approximation to the current number of households in each province. Thus re-weighting is desirable also in the MPHS part of the LSS97 sample.

To compute the correct weights we need to examine carefully the sources of the various sampling parameters used. Where sampling is with probability proportional to size we must identify these size measures so that the values used can be incorporated into the weights. Where sampling is with equal probability we need to identify the numerator and denominator of the sampling fraction.

The weights to be applied are the inverse of the sampling probabilities. Hence we need to identify these probabilities for each of the three sampling stages (communes, villages, households) and for the subsampling stage applied to the MPHS. The overall weight needed for each household is obtained by multiplying together these three or four sampling probabilities and taking the inverse.

Table 6 identifies the relevant parameters and their sources.

Using the symbols shown in Table 6, we may now write the weights.

For the households derived from the LSS93 communes (whether panel or replacement):

$$W_{pij} = \frac{4800}{6090} \cdot \frac{\Sigma_i N_{o,pi}}{c_p N_{o,pi}} \cdot \frac{\Sigma_j M_{1,pij}}{2 M_{1,pij}} \cdot \frac{M_{2,pij}}{16} \cdot \frac{M_{3,pij}}{M_{2,pij}}$$

$$= \frac{4800}{6090} \cdot \frac{N_{o,p}\ M_{1,pi}\ M_{3,pij}}{32\ c_p\ N_{o,pi}\ M_{1,pij}}$$

This weight needs modifying for the communes of Domain 5 (the domain in which the 30 communes were reduced to 27). For the households in the 27 communes retained, insert the further multiplier 30/27 into the weight. The 96 interviews that were rejected were replaced by 90 more from the MPHS, since there were supposed to be about 6000 interviews in all. With this arrangement, the weighted sample size becomes 4800+1290 = 6090 although the actual number of interviews is 4800-96+1290 = 5994. When we put together the two samples they should be weighted according to their sample size: thus the factors 4800/6090 and 1290/6090 appear at the start of the weighting formulas.

For the households of the MPHI subsample the weights are:

$$W_{pij} = \frac{1290}{6090} \cdot \frac{\Sigma_i N_{o',pi}}{c_p N_{o,pi}} \cdot \frac{c_{p'}}{c_p''} \cdot \frac{\Sigma_j N_{1,pij}}{2N_{1,pij}} \cdot \frac{M_{2,pij}}{15} \cdot \frac{M_{3,pij}}{M_{2,pij}}$$

$$= \frac{1290}{6090} \cdot \frac{1}{30} \cdot \frac{c_{p'}}{c_p c_p''} \cdot \frac{N_{o',p}}{N_{o',pi}} \cdot \frac{N_{1,pi}}{N_{1,pij}} \cdot M_{3,pij}$$

Any household observation y_{pijk} contributes to the estimate of the total Y through the formula

$$Y = \Sigma_p \Sigma_i \Sigma_j \Sigma_k (W_{pij} y_{pijk})$$

Thus, to estimate the total we only have to multiply each observation by its appropriate weight and form the sum over all households in the complete sample (both parts).[4]

Note that once the weights have been computed and applied there is no further need to distinguish the two samples.

[4] Since all households in one selected village/precinct are selected with the same probability, the subscript k, for household, does not appear in the weighting formula.

TABLE 6. Sources of sampling parameters[5]

Parameter	Source	Symbol
LSS93		
1. MOS for selection of communes	No. households, 1989 census	$M_{o,pi}$
2. MOS for selection of villages/precincts	No. households as reported by commune 1992	$M_{1,pij}$
3. Frame size, households	Survey team's household listing, 1993	$M_{2,pij}$
4. Frame size, households	Survey team's household listing, 1997/98	$M_{3,pij}$
5. No. communes selected	See list, Survey report	c_p
MPHS		
6. MOS for selection of communes	1989 census pop. updated at province level in 1993	$N'_{o,pi}$
7. MOS for selection of villages/precincts	Pop. reported by commune in sample villages, 1994	$N_{1,pij}$
8. Frame size, households	Survey team's household listing, 1995	$M_{2,pij}$
9. Frame size, households	Survey team's up-date household listing, 1997/98	$M_{3,pij}$
10. No. communes selected per city/town/province	Whole sample: Tab.6, cols 5,6,7 Excl. overlap: cols.8,9,10 Subsample: cols.14,15,16	c_p c'_p c''_p

[5] Symbols: M: households N: Population
Subscripts: p=province; i=commune; j=village/precinct; k=household.
Measures of size for PPS selection: MOS
Number of units listed: Frame size

8. DISTRIBUTION OF THE SAMPLE OVER THE SURVEY MONTHS

Once the sample of communes has been selected a timetable of field work has to be prepared: the work in each commune has to be allocated to one of the teams and one of the months.

The allocation between months should be arranged as far as possible to avoid consistent trends resulting from the allocation. For example, one should not arrange to survey all the cities first, then the towns, then the rural. One should avoid starting systematically with the north and moving steadily southwards over the year. Ideally, every month should be a balanced national sample. This may conflict with organizational needs, but at least every 3-month period should be a well balanced sample. For the same reason, as far as practical it would be desirable to spread any leave periods (rest periods) across the year so that different teams go on leave at different times.