
Nepal Living Standards Survey – III

2010

Sampling design and implementation

Background

The Central Bureau of Statistics (CBS), Government of Nepal, undertook the third Nepal Living Standards Survey (NLSS III) during 2010. The last national living standards survey prior to this, NLSS II, was conducted in 2003-4 and was the basis for the Nepal poverty assessment (World Bank, Resilience Amidst Conflict, 2006). With the NLSS-II survey being about six years old, it is timely that the NLSS-III has been undertaken. This survey, NLSS III, also comes at an important juncture of Nepal's development. Many significant political and economic changes over the last few years have reshaped the country context, and in this evolving environment it is important to update our knowledge of poverty and living conditions of the Nepalese population. The need for such an update is further accentuated by the current global financial crisis the effects of which are also being felt in Nepal.

The NLSS III will visit two independent samples of households:

- A cross-sectional sample with a nominal size of 6,000 households, and
- A panel sample of approximately 1,200 households, previously interviewed in one or both of the previous NLSS-I or NLSS-II surveys.

The rest of this section gives the details of each of the components.

Cross sectional sample

The nominal sample size of the cross-sectional component is 6,000 households. It was selected in three stages:

1. To take advantage of the cartographic segmentation and household listing operations conducted by the CBS for the National Labor Force Survey (NLFS) in 2007/2008, the first sampling stage of the NLSS-3 is identical to that of the NLFS: Using the list of wards and sub-wards identified by the 2000 Population Census as a sample frame¹, the NLFS selected a sample of 800 Area Units, (AUs) allocated into five strata as follows:
 - Mountains: 41 AUs;
 - Urban areas of the Kathmandu valley: 131 AUs;

¹ The list of AUs used as a first-stage sample frame was generated in 2003 as a part of the preparations for the NLSS-2. Most of the AUs are either individual wards or sub-wards, although groups of contiguous wards or sub-wards were occasionally considered as single AUs, to ensure that each had at least 30 households. Some of the larger AUs were later segmented by the NLSS-2, or by the NLFS itself.

- Other urban areas in the hills: 92 AUs;
- Rural hills: 179 AUs;
- Urban terai: 170 AUs; and
- Rural terai: 187 AUs.

Within each stratum, the AUs were selected with probability proportional to size (*pps*) using the number of households as a measure of size (*mos*) and implicit stratification by district.

2. In the second stage, 500 of the NLFS AUs were selected with the following explicit sub-stratification²:

- Mountains: 34 AUs;
- Urban areas of the Kathmandu valley: 72 AUs;
- Other urban areas in the hills: 40 AUs;
- Rural hills East: 32 AUs (*);
- Rural hills Central: 40 AUs;
- Rural hills West: 40 AUs;
- Rural hills Mid-West: 28 AUs (*);
- Rural hills Far-West: 15 AUs;
- Urban terai: 56 AUs;
- Rural terai East: 40 AUs;
- Rural terai Central: 40 AUs;
- Rural terai West: 29 AUs (*);
- Rural terai Mid-West: 20 AUs (*);
- Rural terai Far-West: 14 AUs (*);

Within each sub-stratum, the sample was selected with equal probability, and it was implicitly stratified by district.

² The splitting of the rural hills and rural terai into explicit sub-strata intends to get from the NLSS-3 better estimations by development region than those that could be given by its predecessors. The sub-strata flagged with a star (*) are those where the NLSS-3 will visit all of the NLFS AUs.

3. In the third stage, using the computerized household listings generated by the NLFS as a sample frame, 12 households were selected with equal probability in each AU.³

Panel sample

The panel sample of the NLSS-3 is composed of all households visited by the NLSS-2 in 100 of its primary sampling units (PSUs.) Fifty of them were taken from the cross-sectional component of the NLSS-2, and the remaining fifty from its panel component. In other words (and ignoring attrition,) one half of the NLSS-3 panel households will have been visited for the first time by the NLSS-2, whereas the other half will have been visited by both the NLSS-1 and the NLSS-2.

Analytic domains

Many times the strata and the analytical domains are the same, but some other times they are slightly different, such as in the case of the NLSS III. Due to insufficient sample size in some of the 14 strata for required statistical precision, the strata are not good candidates for use as analytical domains. Instead, estimates can be provided for the 12 analytical domains shown in column 3 of Table 1 below:

³ An additional 6 households were selected in each AU, to be used as a last-resource reserve for non-response among the 12 originally selected households. The 20 households visited by the NLFS in each AU were neither excluded from nor explicitly included in the NLSS-3 sample.

Table 1: Distribution of the Area Units (Aus) within the NLSS-III's strata and analytic domains and within the NLFS's strata

NLSS-III						NLFS				Total Size					
Strata			Analytic domains			Strata									
Stratum code	Stratum name	Number of AUs	Analytic domain code	NLSS III Analytic domain name	Number of Aus	Sample size	NLFS Stratum code	NLFS Stratum name	NLFS number of Aus		NLFS Nominal Sample Size				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)						
100	Mountains	34	100	Mountains	34	408	1	Mountains	41	820	41	321,680			
218	KTM valley Urban	72	218	KTM valley Urban	72	864	2	KTM valley Urban	131	2,620	131	227,637			
219	Other Hills Urban	40	219	Other Hills Urban	40	480	3	Other Hills Urban	92	1,840	92	161,922			
221	Rural Hills E	32	221	Rural Hills E	32	384	4	Rural hills	179	3,580	55	1,619,440			
222	Rural Hills C	40	222	Rural Hills C	40	480							28		
223	Rural Hills W	40	223	Rural Hills W	40	480								15	
224	Rural Hills M	28	226	Rural Hills M - F	43	516									20
225	Rural Hills F	15													
310	Urban Terai	56	310	Urban Terai	56	672	5	Urban terai	170	3,400	170	294,751			
321	Rural Terai C	40	321	Rural Terai C	40	480	6	Rural terai	187	3,740	29	1,686,317			
322	Rural Terai E	40	322	Rural Terai E	40	480							20		
323	Rural Terai W	29	323	Rural Terai W	29	348								14	
324	Rural Terai M	20	326	Rural Terai M - F	34	408									20
325	Rural Terai F	14													
TOTAL		500			500	6,000			800	16,000	800	4,311,747			

Selection probabilities and sampling weights in the original design

The formulas for the selection probabilities and sampling weights are reproduced here.

The probability p_{hi} of selecting ward h_i in NLFS stratum h of the NLFS is given by:

$$p_{hi} = \frac{k_h \times n_{hi}}{N_h} \quad (1)$$

where:

k_h : is the number of wards selected by the NLFS in NLFS-stratum h

n_{hi} : is the number of households in ward hi as per the sample frame

N_h : is the number of households in NLFS stratum h ; given by column (10) in Table 1.

The probability p_{gi} of selecting ward g_i of NLSS-III substratum h is given by:

$$p_{gi} = p_{hi} \times \frac{\tilde{n}_g}{\tilde{n}'_g} \quad (2)$$

where

p_{hi} : is given by formula (1)

\tilde{n}_g : is the number of wards selected by the NLSS-III in substratum g , given by column (4) in Table 1.

\tilde{n}'_g : is the number of wards selected by the NLFS in substratum g , given by column (9) in Table 1.

The probability p_{gih} of selecting household ghi in ward gi of NLSS-III substratum g is given by:

$$p_{gih} = p_{gi} \times \frac{m_{gi}}{n'_{gi}} \quad (3)$$

where

p_{gi} : is given by formula (2)

m_{gi} : is the number of households visited by the NLSS-III in ward gi (nominally always 12)

n'_{gi} : is the number of households listed by the NLFS in ward gi .

To obtain unbiased estimates from the NLSS-III, survey results need to account for the sampling weights (or raising factors) equal to the inverse of their selection probabilities, given by formulas (1), (2) and (3) above. That is:

$$w1_{hgi} = \frac{N_h \times n'_{gi} \times \tilde{n}'_g}{k_h \times n_{hi} \times \tilde{n}_g \times m_{gi}} \quad (4)$$

where $w1_{hgi}$ is the weight for each of the households in ward gi the individual expressions are as defined above.

Post-stratification adjustment of weights

To account for deviations in implementation from sample design, we adjust the sampling weights from (4) in two ways

Survey weights estimate projected population for survey year

The original sample weights are further scaled so that the estimated population from the survey equals the rural and urban population projection for the survey year 2010. For each household j of ward gi , the population scaled weights is:

$$w2_{hgi} = w1_{hgi} \times \frac{U \times P_U + (1 - U) \times P_R}{U \times \tilde{P}_U + (1 - U) \times \tilde{P}_R} \quad (5)$$

where:

$w1_{hgi}$: is as defined in equation (4)

U : indicates whether ward gi is urban. It takes the value of 1 for urban wards and 0 for rural wards

P_U, P_R : are urban and rural population projections, based on CBS estimates.

\tilde{P}_U, \tilde{P}_R : are the population estimates based on the expansion factor, w_{ghi} for urban and rural areas. More specifically:

$$\begin{aligned} \tilde{P}_U &= \sum_{g,i,j} (w1_{hgi} \times S_{hgi} \times U) \\ \tilde{P}_R &= \sum_{g,i,j} (w1_{hgi} \times S_{hgi} \times (1 - U)) \end{aligned} \quad (6)$$

where S_{hgi} is the size of household gij

Survey weights are rescaled to be ensure seasonal balance

The NLSS-III is a year-long survey with fieldwork explicitly stratified into three seasons of four months each. Wards in each of the NLSS-III strata, g , were randomly split into three parts to achieve seasonal representation of each of the strata. We rescale the weights to explicitly make the survey weights equally distributed in each of the seasons to get the final weights for each household. That is,

$$\begin{aligned} w_{hgi} &= w2_{hgi} \times \frac{P/3}{(\tilde{P}_{T_1} \times T_1 + \tilde{P}_{T_2} \times T_2 + \tilde{P}_{T_3} \times T_3) / \tilde{P}} \\ &= w2_{hgi} \times \frac{P \times \tilde{P}}{3 \times (\tilde{P}_{T_1} \times T_1 + \tilde{P}_{T_2} \times T_2 + \tilde{P}_{T_3} \times T_3)} \end{aligned} \quad (7)$$

where:

$w2_{hgi}$: is as defined in equation (5)

$P = P_U + P_R$: is the total projected population for the survey year 2010

$\tilde{P} = P$: is the total estimated population using weights $w2_{hgi j}$, which is the same as P

T_1, T_2, T_3 : are indicators for if ward gi was interviewed in season 1, 2 or 3 respectively

$\tilde{P}_{T_1}, \tilde{P}_{T_2}, \tilde{P}_{T_3}$: are population estimated by weights $w2_{ghi j}$ for season 1, 2, and 3 respectively.
More specifically:

$$\tilde{P}_{T_k} = \sum_{g,i,j} (w2_{hgi j} \times S_{hgi j} \times T_k) \quad (8)$$

For $k = 1, 2, \text{ and } 3$ representing each of the seasons