10th National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 1-2, 2007

The Development of a Master Sample for Bhutan

by

Dalisay S. Maligalig, Khandu Dorji, Melissa C. Pascua,
Nima Deki Sherpa, Peden, and Dorji Wangdi

For additional information, please contact:

Author's name	:	Dalisay S. Maligalig
Designation	:	Senior Statistician
Affiliation	:	Asian Development Bank
Address	:	6 ADB Avenue, Mandaluyong City, Philippines
Tel. no.	:	(0632) 632-5768
E-mail	:	dmaligalig@adb.org
Co-author's name	:	Khandu Dorji
Designation	:	Senior Statistical Officer
Affiliation	:	National Statistics Bureau
Address	:	Royal Government of Bhutan, P. O. Box No. 338 Thimpu, Bhutan
Tel. no.	:	(975) 2 322753
E-mail	:	khadorji_nsb@yahoo.com

Co-author's name	:	Melissa C. Pascua
Designation	:	Assistant Economics and Statistics Analyst
Affiliation	:	Asian Development Bank
Address	:	6 ADB Avenue, Mandaluyong City, Philippines
Tel. no.	:	(0632) 632 6439
E-mail	:	mpascua@adb.org

Co-author's name	:	Nima Deki Sherpa
Designation	:	Computer Programmer
Affiliation	:	National Statistics Bureau
Address	:	Royal Government of Bhutan, P. O. Box No. 338 Thimpu, Bhutan
Tel. no.	:	(975) 2 322753
E-mail	:	nsbnima@yahoo.com

Co-author's name	:	Peden
Designation	:	Computer Programmer
Affiliation	:	National Statistics Bureau
Address	:	Royal Government of Bhutan, P. O. Box No. 338 Thimpu, Bhutan
Tel. no.	:	(975) 2 322753
E-mail	:	peden_nsb@hotmail.com

Co-author's name	:	Dorji Wangdi
Designation	:	Statistical Investigator
Affiliation	:	National Statistics Bureau
Address	:	Royal Government of Bhutan, P. O. Box No. 338 Thimpu, Bhutan
Tel. no.	:	(975) 2 322753
E-mail	:	dorjiee_nsb2006@hotmail.com

The Development of a Master Sample for Bhutan

by

Dalisay S. Maligalig, Khandu Dorji, Melissa C. Pascua,
Nima Deki Sherpa, Peden and Dorji Wangdi[1]

ABSTRACT

This paper outlines the processes that were undertaken in the development of the master sample, namely: the construction of the sampling frame, the determination of sample size and sample allocation across domains; the stratification of the primary sampling units (PSUs) and the allocation of the sample across PSUs and the selection and implementation procedures of PSUs and households for Bhutan, one of the most isolated but the happiest least developed country in the world that is sandwiched between the People's Republic of China and India. The resulting master sample provides a more efficient set of PSUs with survey weights that do not vary very much within domains. The master sample is also expected to provide equally efficient estimates for the two classes of estimates discussed above – for subgroups that cuts across domains and for estimates at the domain level.

I. Background

The Kingdom of Bhutan is a small landlocked mountainous country in the Himalaya Mountains in between India and the People's Republic of China. It is one of the most isolated but the happiest least developed country in the world. The National Statistical Bureau (NSB), the government agency that is mandated to compile the country's official statistics in Bhutan, is steadily becoming a modern statistical system through the assistance of international organizations like the Asian Development Bank (ADB). NSB has conducted the Household Income and Expenditure Survey (HIES) in 2000, the Bhutan Living Standards Survey (BLSS) in 2003. Other government agencies also conduct surveys of national coverage. For example, the Ministry of Labor and Human Resources has the annual labor force survey; the Ministry of Health conducted the Health Survey in 1994 and 2000; the Ministry of Education -- the Literacy Survey in 2004; and the Ministry of Agriculture – the Annual Agricultural Production Survey. In addition, non-government organizations also conduct large-scale surveys.

Prior to the first population and housing census (PHCB) in 2005, the household surveys that were conducted in Bhutan used varied sampling strategies that mostly employ

[1] Senior Statistician, Asian Development Bank (ADB); Senior Statistical Officer, National Statistics Bureau (NSB), Bhutan; Assistant Economics and Statistics Analyst, ADB; Computer Programmer, Statistical Investigator, and Computer Programmer, NSB, respectively. The views of the authors do not necessarily reflect the views of ADB. The authors acknowledge Director Kuenga Tshering for his support to this endeavor; and Ms. Marissa Barcenas for her data processing assistance. This paper will be presented in the Philippine National Convention on Statistics on 2 October 2007 at Edsa Shangri-la Hotel, Mandaluyong City.

area sampling frames that were constructed on the basis of data provided by key informants. Moreover, there was no common sampling design nor common sampled units and thus, surveys were analyzed independent of each other. The PHCB provided an opportunity to develop a master sample that is basically a sample from which sub samples can be selected to serve the need of more than one survey or survey round. Since the master sample can be used by all future household surveys, the cost of preparing the sampling design can be reduced. Surveys that use the master sample will have common sampling units and hence, they can be linked to further enrich data analysis and broaden research.

The master sample was developed by the authors under an apprenticeship program that was funded by an Asian Development Bank's (ADB) technical assistance project: TA 4424 BHU "Strengthening the National Statistical System" Phase II. The authors benefited from the results and technical documentation of the 2003 Master Sample for the Philippines that was developed for the National Statistics Office, Philippines under TA 3656 PHI: Improving Poverty Monitoring Surveys. The Census and Survey Division of the NSB is responsible for the master sample implementation and will provide all technical assistance to the other agencies regarding this master sample.

This paper outlines the processes that were undertaken in the development of the master sample, namely: the construction of the sampling frame, the determination of sample size and sample allocation across domains; the stratification of the primary sampling units (PSUs) and the allocation of the sample across PSUs and the selection and implementation procedures of PSUs and households.

II. Sampling Frame of PSUs

In practice, the PSUs should be areas that have clear boundaries and should also have data and information that could be used for stratification and for sample selection. Because the master sample will be used by many surveys, the PSUs have to be sufficient in size to enable the rotation of sampled households so that response burden of households can be reduced.

Bhutan has three administrative regions – Central, Eastern, Western. Dzongkhags or what is equivalent to provinces in the Philippines, comprise the administrative regions. Similarly, the dzongkhag is composed of gewogs or towns. Rural gewogs consist of chiwogs or villages while Urban gewogs consist of urban blocks.

PSUs can be some level of administrative unit, such as dzongkhags or gewogs. However, all dzongkhags have to be represented in the sample and hence, drawing only a few of them will not be workable. On the other hand, there are 485 gewogs, but they are still considered few to be a good choice for PSUs. This conjecture will be further supported by discussions on sample allocation. Table 1 shows the distribution of households in regions and dzhongkhags in Bhutan.

Table 1 :Summary statistics of total households of Bhutan by Dzongkhag.

Region	Dzongkhag	No. of gewogs/ urban blocks	Households				
			Total	Min	Median	Mean	Max
Central	BUMTHANG	10	2870	47	181	287	798
	DAGANA	17	3485	29	217	205	464
	SARPANG	35	8211	36	129	235	784
	TRONGSA	6	2739	220	499	457	556
	TSIRANG	14	3651	148	232	261	529
	WANGDUE	21	6227	66	297	297	545
	ZHEMGANG	15	3379	42	155	225	597
Eastern	LHUENTSE	11	3001	40	243	273	490
	MONGGAR	29	7348	33	227	253	710
	PEMAGATSHEL	16	2937	6	79	184	720
	SAMDRUPJONGKHAR	39	8363	2	89	214	1292
	TRASHIGANG	29	10813	33	391	373	1064
	TRASHIYANGTSE	14	3764	19	307	269	671
Western	CHHUKHA	72	14482	2	94	201	1431
	GASA	5	727	48	169	145	229
	HAA	9	2290	39	167	254	660
	PARO	15	7118	42	424	475	1341
	PUNAKHA	10	3387	124	306	339	816
	SAMTSE	36	11634	1	226	323	1018
Thimphu	THIMPHU Rural	10	3961	37	351	396	844
	THIMPU City	72	15728	2	166	218	978
Bhutan		485	126115	1	184	260	1431

Source: PHCB 2005.

As shown in Table 1, the number of households in gewogs varies widely, from only 1 to 1431. This wide variation in the number of households will result to wide variation in the

selection probabilities of households, which is not preferred because as a consequence, there will be large variations in survey weights that will contribute to a large increase in the variance of the estimate. Probability proportional sampling at the PSU selection stage can address this problem only to a certain extent. However, other controls such as dividing large gewogs and collapsing small adjacent ones have to be done in order that the final PSU composition or clusters will not vary widely. Also, to avoid multiple selection, large PSUs have to split into smaller PSUs so that every PSU will have only one chance of inclusion in the sample. Similarly, those smaller PSUs size have to be merged to the adjacent PSUs. In so doing, a better approach would be to designate the chiwog (villages) or in case of urban gewogs, urban blocks or enumeration areas (EAs) as delineated in the 2005 PHCB, as the PSUs. To ensure that there will be sufficient number of households for rotation purposes, chiwogs/EAs that have less than 100 households are combined with an adjacent chiwog. The summary of households by chiwog/EAs is in Table 2.

Table 2: Summary Statistics of Total Households of Bhutan by Dzongkhag

Region	Dzongkhag	Chiwog/ EAs	Households				
			Total	Min	Median	Mean	Max
Central	BUMTHANG	20	2870	108	134	144	223
	DAGANA	21	3485	104	146	166	532
	SARPANG	41	8211	105	175	200	397
	TRONGSA	15	2739	122	185	183	273
	TSIRANG	13	3651	148	259	281	529
	WANGDUE	29	6227	106	172	215	545
	ZHEMGANG	18	3379	107	158	188	325
Eastern	LHUENTSE	18	3001	109	131	167	391
	MONGGAR	34	7348	109	175	216	384
	PEMAGATSHEL	15	2937	109	169	196	367
	SAMDRUPJONGKHAR	55	8363	102	132	152	377
	TRASHIGANG	74	10813	103	130	146	391
	TRASHIYANGTSE	20	3764	109	137	188	390

Western	CHHUKHA	78	14482	113	167	186	369
	GASA	4	727	132	183	182	229
	HAA	13	2290	102	160	176	324
	PARO	43	7118	100	146	166	424
	PUNAKHA	15	3387	111	227	226	368
	SAMTSE	69	11634	102	149	169	384
Thimphu	THIMPHU Rural	19	3961	108	193	208	338
	THIMPHU City	81	15728	106	182	194	377
Bhutan		695	126115	100	156	181	545

Source: PHCB 2005; Authors' computations.

III. Sample Size and Sample Allocation

What are the domains?

The sample size is usually determined at the level from which separate estimates will be derived. This level is called the domain – the designated subpopulations for which separate samples are planned, designed and selected (Kish,1987). Hence, the domains have to be designated first before the total sample size can be determined. The sample that will be drawn at the domain level will be such that it can produce reliable estimates for any type of surveys in the country.

The domains in the past surveys are usually the administrative regions –Central, Eastern, Western. However, as Table 3 and also the results of past surveys show, Thimphu City's characteristics are different from the administrative regions. For example, Table 3 shows that Thimphu has 17.1% poverty rate, significantly less than the three other regions. Also, about 12.5% of the households live in Thimphu City, making it the largest city/dzhongkhag in Bhutan. Because of these reasons, Thimphu City was also included as a domain for the purposes of the master sample. It is also interesting to note that in the 2003 BLSS, Thimphu City is also considered a domain.

Table 3: Poverty rates with measures of precision, by region

Region	Rate (%)	SE(%)	CV(%)
Central	52.3	2.9	5.6
Eastern	69.5	3.7	5.3
Western	43.0	5.5	12.8
Thimphu city	17.1	3.2	18.8

Source: BLSS 2003.

What is the total sample size?

The sample size for a domain is usually determined based on the variability of the sampling units within the domain, the error level that is acceptable and the associated costs. An optimum number of households to select from each sample PSU are first determined and then the number of PSUs is computed. The following is the description of choosing the domains, how the total sample size was determined and allocated across the domains and how the number of households is to be selected in each PSU.

Table 3 shows the estimated poverty rates for the four regions with their estimated standard errors (SEs) and coefficient of variations (CVs). The CVs are inversely related to the size of percentage estimates. For example, the high CV of 18.8 percent for the Thimphu city estimates of 3.2 percent in poverty. This CV is high because the percentage is also low. The SE of this percentage is 3.2 percent, translating into a 95 percent confidence interval for the Thimphu City percentage in poverty from 10.8 to 23.4 percent.

To determine the total sample size, the tentative sample size, if the sampling design is simple random sampling, is inflated by the design effect of a major characteristic from existing surveys. The design effect (*deff*) is the ratio of the variance of the estimate of a major characteristic taken from the previous survey to the variance of the estimate on the assumption that the survey has a simple random sample design. The design effect is a measure of the efficiency of the design of the survey and it provides a useful mechanism in designing future surveys.

The two existing surveys from which the design effects were computed are the 2003 BLSS and the 2000 HIES. These surveys, however, are based on a sampling design that resulted in a wide variation in weights within domains. In the 2003 BLSS for example, PSUs were allocated equally with 60 for each urban domain and 30 for each rural domain. Within each domain, the number of PSUs for a dzhongkhag was not determined on the basis of the total number of households, hence, the resulting selection probabilities were not uniform within domains. More so, when the listing operation of the survey resulted in a different number of households than what was used in the computation of weights, instead of retaining the original survey weight for the PSU and adjusting the number of households to be sampled, what was done was the opposite; thus adding more variation to the survey weights. Table 4 supports this discussion further.

Table 4. Survey Weights for Western Region Rural Gewogs, 2003

Name of Gewog	Total Households 1999	Total Households 2003	Sample Size	Survey Weight
Bapisa	301	380	20	54.28
Genye	144	149	20	44.48
Mewang	577	501	20	37.329
Lingzhi	100	99	20	42.562
Dopshari	340	399	17	59.355
Lango	496	495	19	45.163
Naji	281	488	20	74.662
Wangchang	437	350	19	36.245
Bji	440	408	20	39.865
Sangbay	246	230	16	50.244
Charghare	655	679	20	44.567
Chengmari	549	814	20	63.744
Nainital	317	392	20	53.163
Pagli	757	705	19	42.146
Tading	966	864	20	38.452
Dorokha	802	779	19	43.957
Mayona	326	230	20	30.332
Biru	778	616	20	34.04
Lahireni	415	498	20	51.59
Tendru	618	645	19	47.232
Bjabchho	223	137	20	26.412
Bongo	758	695	20	39.418
Dungna	292	275	20	40.489
Dala	916	928	20	43.555
Logchina	434	450	19	46.923
Phuntsholing	870	851	20	42.053
Guma	334	307	19	41.60
Kabjisa	466	411	20	37.918
Teowang	317	219	19	31.264
Goenkhatoe	89	52	19	26.441
Total	14244	14046	584	

Source: BLSS 2003.

The wide variation in weights that was shown in Table 4 will certainly result in bigger design effects that will in turn unnecessarily increase the sample size. Hence, the contributions of the variation of weights to the design effects have to be taken out. To do this, the increase in variance due to the variation of weights, *(1+L),* has to be estimated. Kish (1992) estimates this in terms of relative variance as: $1+L=1+cv(w)^2$, where $cv(w)$ is the coefficient of variation of weights and in this case, can computed from the 2003 BLSS. Thus, the adjusted design effect deff_{ms} will be:

$$\text{deff}_{ms}=\frac{\text{deff}_{\text{BLSS}}}{1+L}=\frac{\text{deff}_{\text{BLSS}}}{1+cv(w)^2}$$

Table 5 presents the basic statistics on the survey weights in each domain, the increase in variance, the original and adjusted design effects.

Table 5: Adjustment of Design Effects Due to Variation in Survey Weights

Region	Weights						Increase in Variance	Deff from poverty rate	Adjusted deff
	No. of hhs	Mean	Std. Dev.	Min	Max	CV (weights)			
Central	1184	21.17	16.08	3.13	54.87	0.7596	1.58	3.54	2.25
Eastern	1077	35.53	33.66	2.69	94.74	0.9473	1.90	8.76	4.61
Western	1093	26.79	16.89	6.63	74.66	0.6303	1.40	14.16	10.13
Thimphu	653	21.84	9.35	9.38	54.28	0.4283	1.18	3.42	2.89

Source: Authors' computations.

The first iteration in the determination of sample size assumes that given the sample estimate *p* of the poverty rate, a margin of error *d=0.05* and risk =0.05 that we are willing to incur, the tentative sample size for a domain, assuming simple random sampling is

$$n_{SRS} = \frac{\frac{t^2_{(a,N-1)}PQ}{d^2}}{1+\frac{1}{N}\left(\frac{t^2_{(a,N-1)}PQ}{d^2}-1\right)},$$

where $t_{(a,N-1)}$ is the abscissa of the t-distribution given a and the population size *N*; *P* is the true proportion and *Q=(1-P).* This sample size is then inflated by the adjusted design effect from table 5, such that the tentative sample size for a complex survey will be:

$$n_{complex} = Deff \times n_{SRS}$$

Table 6: Tentative Sample Sizes Using Adjusted Design Effect At 5% Error Level

REGION	Poverty Rate (%)	Total households	Error level	Tentative Sample Size	SRS sample size	DEFF	Tentative sample size
			0.05				
Central	52.02	30562	1.96	170.47	169.53	2.25	380.73
Eastern	69.55	36226	1.96	144.65	144.08	4.61	664.85
Western	43.04	39638	1.96	167.43	166.73	10.13	1689.67
Thimphu	17.12	19689	1.96	96.92	96.45	2.89	278.33
		126115					3013.58

Source: Authors' computations.

How is the total sample size allocated across domains?

The tentative sample sizes in Table 6 vary widely with the Western region taking the lion share of more than 50% of the sample. While the total of about 3,000 is acceptable and would present a significant reduction from the total sample size of the previous survey, 2003 BLSS, it is noted that there is a price to pay for over sampling some domains in that the sampling errors of other estimates that are not domain specific are increased. For example, over sampling of small geographic domains will lead to less precision for estimates for socio-demographic subgroups that cut across the geographic domains. In practice, survey estimates are needed for the country as a whole as well as for various subgroups. For example, these subgroups include socio-demographic subdivisions that are usually spread throughout the country, such as female-headed household by age of head, or educational levels by age and sex. They also include geographic subdivisions, such as dzongkhags.

It has, therefore, become imperative to take a careful look at the sample allocation across domains. Several existing sample allocations were examined on the basis of the expected variability of the estimates, namely:

Equal Allocation: $n_d = \frac{n}{D} = \frac{n}{4}$

Proportional Allocation: $n_d = n\frac{N_d}{N} = nW_d$

Square Root Allocation: $n_d = n\frac{\sqrt{N_d}}{\sum_d \sqrt{N_d}}$

Kish Allocation: $n_d = n\frac{\sqrt{D^{-2} + IW_d^2}}{\sum_d \sqrt{D^{-2} + IW_d^2}} = n\frac{\sqrt{\frac{1}{16} + IW_d^2}}{\sum_d \sqrt{\frac{1}{16} + IW_d^2}}$

where n_d is the sample size in the domain, n is the sample size, D is the number of domains, N_d is the total number of households in domain d, N is the total number of households in Bhutan, per the 2005 Census of Population and W_d is the proportion of households in domain d, and I is the Kish allocation index denoting the relative importance assigned to estimates at the national or subgroups that cut across domains (i) as compared

to those estimates at the domain levels (ii). To illustrate, examples of (i) are: numbers of crop farmers and female headed households, proportions of persons in poverty in Bhutan, number of persons in the labor force who are unemployed, proportion of households with electricity, proportion of women who gave birth with a health personnel in attendance, estimates of the differences between subgroups.

In general, the best approach for estimates in class (i) is to allocate the sample size proportional to the population size of each domain. However, the best approach for estimates in class (ii) is to divide the total sample size equally between the regions. The difference between these two allocations can be substantial when the domains differ in measure of size. If either one of these allocations is used, it will perform well for its class of estimates but perform badly for the other class. A compromise allocation that is suboptimal for both classes of estimates but that performs reasonably well for both is often the preferred solution. This is the Kish allocation. With $I = 0$, the Kish allocation reduces to the equal allocation. With $I \to \infty$, the Kish allocation tends to the proportional allocation.

Table 7: Summary of Results of Various Allocation Methods

Domain (*d*)	N_d	W_d	Equal	Propor-tional	Square Root $\frac{\sqrt{N_d}}{\sum_d \sqrt{N_d}}$	n_d	KISH ALLOCATION (I) 0.25	0.5	0.75	1	2	3
Central	30562	0.24	750	727	174.82	744	742	737	735	733	729	728
Eastern	36226	0.29	750	862	190.33	810	770	784	794	802	820	830
Western	39638	0.31	750	943	199.09	848	789	814	832	845	877	893
Thimphu	19689	0.16	750	468	140.32	597	699	665	640	620	574	549
Bhutan	126115	1.00	3000	3000	704.56	2999	3000	3000	3001	3001	3000	3000

Source: Authors' computations.

The expected variability for each of the allocation methods described above was derived using the following procedures:

1. The estimates of poverty rate, *p*, and design effect by domain were computed from the BLSS, d^2. The estimated design effect is the ratio of the estimated variance of the estimated proportion under the complex sample design to the variance that would have been obtained had the estimated proportion been based on a simple random sample of the same size. This estimated design effect was automatically generated from Stata, with the sampling design characteristics of the BLSS applied; that is, each responding household was identified by domain, area (urban/rural), dzongkhag and PSU.

2. For each domain, $d_w^2 = 1 + cv(w)^2$, where $cv(w)^2$ is the coefficient of variation of the weights for sampled households in a particular domain was computed. d_w^2 is the measure of the increase in variance due to the variation in weights. This procedure was also done in computing for the effective sample size by domain since the existing surveys have weights that vary at the PSU level (refer to section 3).

3. The design effect from step 1 by $d_c^2 = \frac{d^2}{d_w^2}$ was adjusted.

4. The estimated intra-class correlation as: $\hat{r} = \frac{d_c^2 - 1}{b_{BLSS} - 1}$, where b_{BLSS} is the number of households sampled per PSU from the BLSS was computed.

5. The expected design effect, d_e^2, as: $d_e^2 = 1 + (b_{MS} - 1)\hat{r}$, where b_{MS} is the number of households that will be sampled per PSU for the master sample, was computed. In this case, b_{MS} equal to 10, 15, and 20 were tested.

5. The expected CV values by domain were computed as: $cv(p)_{MS} = \sqrt{\frac{d_e^2(1-p)}{n_d\, p}}$, where n_d is the sample size for domain *d* using a particular allocation scheme.

Table 8. Expected Coefficient of Variation (%) for Various Allocation Methods and Sample Sizes (n=10, 15, 20)

Region	Equal Allocation			Proportional			Square Root		
	10	15	20	10	15	20	10	15	20
Central	4.42	4.86	5.26	5.60	6.15	6.65	4.96	5.44	5.89
Eastern	3.98	4.62	5.19	3.55	4.12	4.63	3.74	4.35	4.88
Western	9.69	11.68	13.37	9.04	10.89	12.47	9.33	11.24	12.87
Thimphu	11.05	12.42	13.65	11.23	12.61	13.86	11.10	12.47	13.70
	Kish Allocation								
	I=1			I=2			I=3		
	10	15	20	10	15	20	10	15	20
Central	4.86	5.34	5.78	5.54	6.08	6.58	5.67	6.22	6.73
Eastern	3.75	4.36	4.89	4.03	4.68	5.26	4.00	4.64	5.21
Western	9.38	11.29	12.93	10.16	12.24	14.01	10.09	12.16	13.92
Thimphu	11.18	12.56	13.80	12.28	13.79	15.16	12.29	13.80	15.17

Source: Authors' computations.

On the basis of Table 8, the number of households to be sampled per PSU should be 10 as opposed to 20 in the rural areas in the 2003 BLSS because it gives the least expected coefficient of variation. In fact, the number of households can still be reduced on the basis of

the intra-class correlations that were obtained from the procedure described above. However, it was deemed to fix the number of households at 10 per PSU to cover non-response, if any. Moreover, it is more practical for NSB to have a uniform number of households for all domains.

For the PSU allocation, it seems that there is not much difference among the Kish Allocation at I=1, equal and square root allocations. The final choice can be any of these allocations; however, NSB staff chose the Kish Allocation because of the rationale of the method.

The number of PSUs to be sampled in each domain is then computed by dividing the allocated sample size by the desired sub-sample size by PSUs. This is shown in Table 9. Note that final number of PSUs may differ from the tentative number of PSUs because of the stratification of PSUs within a domain.

Table 9: Distribution of PSUs, SSUs, sampling fraction and weights by region

Region	Total no. of Households	Total no. of sample Households	Selection Probability	Tentative Weights	Tentative Number of PSUs
Central	30562	733	0.023984	41.69441	73.3
Eastern	36226	802	0.022139	45.16958	80.2
Western	39638	845	0.021318	46.90888	84.5
Thimphu	19689	620	0.03149	31.75645	62.0
Total	126115	3000			300.0

Source: Authors' computations.

IV. PSU Stratification

PSUs within a domain have to be stratified to improve the precision of survey estimates and to control costs and administration of the survey. Since NSB has field offices in every dzhongkhag, the dzhongkhag was a good choice as a stratification variable. Also, since the characteristics of households in the urban EAs are very different from those rural chiwogs, then urban/rural classification was also introduced as a stratification measure. In cases of small dzongkhag, in which further stratification is not possible anymore, then urban/rural classification is not introduced; hence. urban EAs are in the same stratum as those rural chiwogs.

To get the tentative number of sample households in a stratum, the selection probability for that domain found in Table 9 is applied in the total number of households in the stratum, the result of which is divided by 10, the number of households per PSU to get the tentative number of PSUs per stratum. This number, however, is rounded to the nearest integer to get the final number of PSUs.

Table 10: Sample size and number of sample PSUs by Region-Dzhongkhag-Area

Region	Dzongkhag	Area	Total Households	No. of sample hhs	Tentative No. of sample PSUs	Final No. of PSUs
Central	BUMTHANG	Rural	2130	51	5.1	5
Central	BUMTHANG	Urban	740	18	1.8	2
Central	DAGANA	Combined	3485	84	8.4	8
Central	SARPANG	Rural	5829	140	14	14
Central	SARPANG	Urban	2382	57	5.7	6
Central	TRONGSA	Rural	2211	53	5.3	6
Central	TRONGSA	Urban	528	13	1.3	1
Central	TSIRANG	Combined	3651	88	8.8	9
Central	WANGDUE	Rural	4937	118	11.8	12
Central	WANGDUE	Urban	1290	31	3.1	3
Central	ZHEMGANG	Rural	2891	69	6.9	6
Central	ZHEMGANG	Urban	488	12	1.2	1
Total			**30562**	**734**	73.4	**73**
Eastern	LHUENTSE	Combined	3001	66	6.6	7
Eastern	MONGGAR	Rural	6187	137	13.7	14
Eastern	MONGGAR	Urban	1161	26	2.6	3
Eastern	PEMAGATSHEL	Combined	2937	65	6.5	7
Eastern	SAMDRUPJONGKHAR	Rural	6214	138	13.8	14
Eastern	SAMDRUPJONGKHAR	Urban	2149	48	4.8	5
Eastern	TRASHIGANG	Rural	9687	214	21.4	21
Eastern	TRASHIGANG	Urban	1126	25	2.5	3
Eastern	TRASHIYANGTSE	Rural	3223	71	7.1	7
Eastern	TRASHIYANGTSE	Urban	541	12	1.2	1
Total			**36226**	**802**	**80.2**	**82**
Western	CHHUKHA	Rural	7690	164	16.4	16
Western	CHHUKHA	Urban	6792	145	14.5	15
Western	GASA	Rural	727	15	1.5	2
Western	HAA	Combined	2290	49	4.9	5
Western	PARO	Rural	6552	140	14	14
Western	PARO	Urban	566	12	1.2	1
Western	PUNAKHA	Combined	3387	72	7.2	7
Western	SAMTSE	Rural	9418	201	20.1	21
Western	SAMTSE	Urban	2216	47	4.7	5
Total			**39638**	**845**	**84.5**	**86**
Thimphu	THIMPHU	Rural	3961	125	12.5	13
Thimphu	THIMPHU	Urban	15728	495	49.5	50
Total			**19689**	**620**	**62**	**63**

Source: Authors' computations.

V. Sample Selection

To reduce the increase in variance due to large variation in weights, the master sample design calls for all households in a given domain to have the same overall selection

probability. This section discusses the selection probabilities and implementation procedures needed to achieve this objective.

The plan calls for two stages of sampling in domain *d*: (i) PSU a will be selected and (ii) household b in a PSU a will be selected. Before selecting the PSU, however, stratification will be introduced: (i) by dzongkhag; (ii) and if the dzongkhag has large number of psus, then by rural/urban areas. Table 10 lists all the strata by domain and dzongkhag.

The probability f_d for domain *d*. that a household is selected is given by:

$$f_d = P(ha)P(hb|a) \tag{1}$$

where $P(ha)$ is the selection probability of PSU a, and $P(hb|a)$ is the selection probability for household b given in PSU a is sampled, all in stratum *h*.

$$f_d = P(hab) = \frac{a_h M_{ha}}{\sum_a M_{ha}} \frac{b_h}{M_{ha}} = \frac{a_h b_h}{\sum_a M_{ha}} \tag{2}$$

In this equation, $P(hab)$ is the selection probability for household b in PSU a in stratum *h*, M_{ha} is the measure of size for PSU a in stratum *h*, a_h is the number of PSUs to be sampled from stratum *h*, and b_h is a number of households to be selected from stratum h that has to be determined to satisfy the equation (1).

The term $P(hb|a) = \frac{b_h}{M_{ha}}$ in equation (2) represents the sampling fraction to be used in the systematic sampling of housing units at the final sampling stage. Its inverse is the sampling interval to be applied.

Because the nonresponse in both BLSS and HIES is quite nil, non-response adjustments to the sample size was not introduced in planning the sampling strategy. Also, since the total population count of Bhutan is much smaller than the projected population count, non-coverage adjustment was not also introduced. Moreover, the Census of Population, which is the basis of the sampling frame of PSUs has just been conducted in 2005; hence the need for non-coverage adjustment is not immediate.

Implementation Procedures

How to Select PSUs

This subsection describes the steps implemented on the sampling procedures for PSUs described above. The steps of the sampling procedure were as follows:

1. The number of PSUs to be selected in stratum *h,* a_h, was determined (see Table 10). Basically,

   $a_h = \frac{n_h}{b}$, where *b* is the recommended number of households per PSU determined in Section 3, while n_h is the number of households allocated to stratum *h*.

2. The value of the actual number of households, b_h to be drawn was derived as:

$$b_h' = \frac{f_d \sum_a M_{ha}}{a_h}$$

   where f_d is the sampling fraction for domain *d*. b_h is the rounded figure of b_h'.

3. The final sampling weight for stratum *h* was computed such that

$$w_h = \frac{\sum_a M_{ha}}{a_h b_h}$$

   Note that because b_h' is not exactly equal to b_h, $w_h \neq \frac{1}{f_d}$ but very close.

4. The sampling interval, $S = \frac{M_{ha}}{b_h}$, was computed.

5. A random start (RS) was selected by drawing a random number between 0 and 1 and multiplying it by the interval in step 4. The first sampled PSU was the first PSU with cumulative value of M_{ha} containing the value of the random start (RS). The next sample PSU was the PSU for which the cumulative value of M_{ha} contains RS+*S*, the next was the PSU for which the cumulative value contains RS +*2S*, etc.

Drawing Households from a Sampled PSU

There are two approaches to selecting households from a sampled PSU:

(i) Selection can be done randomly from the list of households in the 2005 CPHB; or

(ii) Selection can be done after a household in a selected PSU has be listed again.

Doing item (i) will not entail additional cost in survey operations. However, the list of households in the 2005 CHPB will become obsolete as the years progressed. Hence, it becomes imperative to conduct listing operations in selected PSUs after several years.

In the case of the second approach, if the total number of households may not be the same anymore as that from the 2005 Census of Population, to maintain the same household selection probability in a particular stratum, the number of households to be sampled, b_h', has to be adjusted such that:

$$b_h' = b_h \frac{M_{ha}'}{M_{ha}}$$

where M_{ha}' is the updated total number of households in PSU a based on the listing operations.

There are 304 PSUs that were drawn in the final list which is found in the Master Sample technical documentation, together with the weighting and estimation procedures that will not be discussed in this paper.

VI. Conclusions and Recommendations

The master sample consists only of 304 PSUs with 3025 sampled households, which, compared to the 2003 BLSS which has 326 PSUs and 4120 households, will reduce survey costs and render smaller sampling errors. As shown in Table 11 below, the master sample provides a more efficient set of PSUs with survey weights that do not vary very much within domains. The master sample is also expected to provide equally efficient estimates for the two classes of estimates discussed above – for subgroups that cuts across domains and for estimates at the domain level.

Table 11. Comparison of Increase in Variance Due to Weights, Master Sample and 2003 BLSS

Domain	No. of PSUs	Statistics on the Survey Weights, Master Sample				Increase in Variance due Weights	
		Mean	Minimum	Maximum	Coefficient of Variation	Master Sample	2003 BLSS
Central	73	41.54	39.44	44.00	0.10	1.01	1.58
Eastern	82	45.55	42.98	46.95	0.08	1.01	1.90
Western	86	47.69	44.86	49.38	0.10	1.01	1.40
Thimphu	63	31.28	30.60	31.91	0.05	1.00	1.18

Source: Authors' computations.

A more efficient approach to selecting PSUs within a domain which will result into a more uniform survey weight will be to draw PSUs systematically with a random start on the full list of PSUs that are arranged according to dzhongkhag and urban/rural classification. A sampling interval that is equal to the total number of households in the domain divided by the number of PSUs to be sampled will be applied to the cumulative total of households. This approach will maintain almost the same selection probabilities of households within a domain. However, it cannot guarantee that small dzongkhags will be represented.

References

Kalton, Graham. 1983. Introduction to Survey Sampling. Sage University Papers.

Kalton, Graham. 2003. Notes on the Master Sample for the Philippines.

Kish, Leslie. 1992. Weighting for Unequal Pi. Journal of Official Statistics, Statistics Sweden. 183 -200.

Kish, Leslie. 1965. Survey Sampling. John Wiley and Sons, Inc.

Kish, Leslie. 1987. Statistical Design for Research. John Wiley and Sons, Inc.

National Statistics Bureau, Bhutan. 2006 Master Sample Documentation

National Statistics Office, Philippines. 2003. 2003 Master Sample Documentation.

Appendix 1

Sampling Design of the 2000 Household Income and Expenditure Survey

Note: This is an excerpt from the Technical Documentation of the 2000 Household Income and Expenditure Survey

In the 2000 HIES there were four stratum. In stratum 1, there were 7 towns and in stratum 2 all the remaining 15 towns. In stratum 3, 22 gewogs each are having at least 750 households and in stratum 4, all the remaining 180 gewogs. The sample of 2000 households for the urban areas having distributed to stratum 1 and stratum 2 in proportion to the number of households. This has resulted in an allocation of 1650 households to stratum 1 and 350 households to stratum 2 . The sample of 1650 households in stratum 1 has being allocated to the 7 towns approximately in proportion to the number of households thus giving a sample of 800 households for Thimphu and 300 household for Phuentsholing, 150 households for Gelephu and 100 households each for Punakha, Samdrupjongkha, Chukha and Wangdue phodrang. In the case of small 15 towns in stratum 2 first they were arranged in descending order of number of households and from this a sample of 7 towns were selected by circular systematic sample. From each of the 7 towns a sample of 50 households was selected. Likewise, for selection of 56 gewogs from stratum 4, all the 108 gewogs were first arranged by Dzongkhag and thereafter a circular systematic sample of 56 gewogs was selected.

Selection of households in towns- Thimphu town has been divided into 70 EAs, each of 100-125 households. A fresh listing of households was carried out under each EA, identifying the households in each house, noting down the name of the head of the households and ascertaining whether the head of the household is an expatriate. All the EAs under Thimphu town were group into 3 socio-economic groups, high, medium and low on the bases of known information about the rent or value of real estate. In this list, all the households of known expatriates were given a running serial number starting with the first household in the first EA to the last household in the last EA. From such a arranged list, a circular systematic sample were selected.

Selection of household in rural areas- In case of rural areas, a sample of 40 households were selected from each of the 22 gewogs in stratum 3 and the remaining sample of 1120 households were selected from stratum 4. For selection for households in each of the 22 gewogs in stratum 3, samples of 4 chiwogs were selected. All households in each selected chiwogs were listed along with the details relating to name of the head of the households, the size of households, prime means of livelihood, (self-employed in non-agriculture; rural labour and others), description of activity of any non-agricultural enterprise operated from within the premises of the households or without any premises along with a broad industry group code. The list of households so prepared was rearranged by PML and by size of the households within each of 3 PML classes. From such an arranged list, circular systematic samples of 10 households were selected for the survey.

The selections of households in stratum 4 were done in 3 stages. As a first step the 180 gewogs were arranged by dzongkhag and from this re-arranged list a sample of 56 gewogs was selected following the procedure of circular systematic sampling. In each selected gewog, 2 chiwogs were selected and from within each chiwog a sample of 10 households was selected.

In 2003 BLSS a stratified two staged sampling design was adopted. There were two primary strata- Urban and Rural. There were 4 regions in the urban and 3 regions in the rural areas. The primary and secondary sample units were blocks/EAs for urban areas and gewogs for rural areas and households respectively.

Town areas: In the urban strata there are 4 regions namely; Thimphu city, West (Thimphu/Rest), Central and East. All dzongkhag headquarters, 5 satellite towns (Bonday, Gedu, Khaling, Rangjung and Wamrong) and 2 dungkhag headquarters namely Phuentsholing and Gelephu towns are also included in the survey. The total number of blocks/EAs was 240 and the total number of sample households was 2400. The sample size is proportional to the total number of households in each town. Within each sample block/EA, a fixed sample of 10 households was selected.

Rural areas: In the rural strata, there were 3 regions namely: West, Central and East. Out of 201 gewogs, 86 sample gewogs were selected by probability proportional to size where size was the

number of households in the gewog. Within each selected gewog, a fixed sample of 20 households was selected.

Region: The regions comprises of the following dzongkhag:
Thimphu: Thimphu city
West: Thimphu rural, Paro, Ha, Samtse, Chukha, Punakha and Gasa
Central: Wangdue Phodrang, Daga, Tsirang, Sarpang, Zhemgang, Trongsa and Bumthang
East: Lhuntse, Monggar, Pemagatshel, Samdrup Jongkhar, Trashigang and Yangtse.

Selection of secondary sampling units: Households are the secondary sampling units. A new listing of households was made for every sample blocks/EAs and gewogs to serve as the sampling frame. Once the listing was completed, the required sample households were selected by circular systematic sampling.

Appendix 2

Sampling Design of the 2003 Bhutan Living Standard Survey

Note: This is an excerpt from the Technical Documentation of the 2003 Bhutan Living Standards Survey

The Bhutan Living Standards Survey followed the Living Standards Measurement Survey (LSMS) methodology developed by researchers at the World Bank. The methodology has been customized for Bhutanese population and collected information through an integrated household questionnaire covering consumption, assets, housing, education, health, fertility, accompanied by a community questionnaire aimed at collecting information on services and prices of the common commodities.

The Survey and Data Processing Division of the National Statistical Bureau maintains an updated list of households by gewog for the entire country. Mapping of the urban centers was conducted and a sketch map was drawn for all the urban centers included in the survey prior to the block selection.

The sample size for the BLSS was determined using the information relating to the coefficient of variation (CV) of some key expenditure variables as recommended by the United Nations Statistical Office based on the log-normal distribution of key expenditure variables. Keeping in view the uses of the survey results, it was felt that sampling precision of 5 percent at the national level would serve the purpose of the survey. Taking the assumption, it was derived that n, the sample size required for estimating the population mean y with 95 percent confidence.

$$n = 1600(CV)^2$$

Taking into account of the CV of 1.0492 for the log-normal distribution and design effect of 2 for a stratified multi-stage sampling design, sample size of 4200 distributed to urban and rural with provisions for some non-response households was determined.

In order to get a representative sample of the households the entire country was divided into three regions based on the number of households and their geographic location; namely Western, Central and Eastern.

Western Region:	Thimphu, Paro, Ha, Samtse, Chhukha, Punakha and Gasa
Central Region:	Wangduephodrang, Daga, Tsirang, Sarpang, Zhemgang, Trongsa and Bumthang
East Region:	Lhuntse, Mongar, Pemagatsel, Samdrup Jongkhar, Trashigang and Trashi Yangtse

The different areas were divided into seven strata. Each stratum composed of four urban strata (Thimphu city, rest of Western, Central and Eastern) and three rural strata (Western, Central and Eastern) has been designed to yield 600 households. Two rural areas of Sarpang and Samdrup Jongkhar were dropped out in the sample due to security reasons. The final sample size for the survey was 4120 households. The distribution of the sample household is shown in Table 1.1.

The urban areas were divided into number of blocks. Each block was formed with a well-defined boundary and number of households ranging from a low of 100 households to maximum of 300 households. Gewogs in the rural area and blocks in the urban area were treated as enumeration areas and thereby the primary sampling units (PSU). The PSUs were selected based on probability proportional to size for both urban and rural areas. Within the selected PSUs a total of 20 households and 10 households were selected based on systematic circular sampling scheme for the Gewogs and blocks respectively.

Table 1.1: Distribution of sample blocks/gewogs and households by Dzongkhag, by Urban and Rural.

Region	Dzongkhag/Town	Urban		Rural		Total
		Sample Blocks	Sample Households	Sample Gewogs	Sample Households	Sample Households
Western	Thimphu	60	600	4	80	680
	Paro	3	30	4	80	110
	Haa	2	20	2	40	60
	Samtse	12	120	10	200	320
	Chhukha	9	90	6	120	210
	Phuentsholing	29	290	-	-	290
	Punakha	4	40	3	60	100
	Gasa	1	10	1	20	30
	Total	120	1200	30	600	1800
Central	Wangduephodrang	14	140	8	160	300
	Dagana	3	30	6	120	150
	Tsirang	5	50	6	120	170
	Sarpang	7	70	Excluded from survey		70
	Gelephug	21	210			210
	Zhemgang	4	40	3	60	100
	Trongsa	3	30	4	80	110
	Bumthang	3	30	3	60	90
	Total	60	600	30	600	1200
Eastern	Lhuntse	4	40	2	40	80
	Mongar	5	50	7	140	190
	Pemagatshel	5	50	3	60	110
	Samdrup Jongkhar	26	260	Excluded from survey		260
	Trashigang	15	150	10	200	350
	Trashi Yangtse	5	50	4	80	130
	Total	60	600	26	520	1120
	Grand total	240	2400	86	1720	4120