Developing a Master Sample Design for Household Surveys in Developing Countries: A Case Study in Bangladesh

Survey Methods: Insights from the Field

Dalisay S. Maligalig | Arturo Martinez Jr.

How to cite this article : Maligalig, D. S., & Martinez, A. Jr (2013). Developing a Master Sample Design for Households Surveys in Developing Countries: A Case Study In Bangladesh. Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=2151

DOI : 10.13094/SMIF-2013-00009

Copyright : © the authors 2013. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Abstract : For evidence-based policy making, socio-economic planners need reliable data to evaluate existing economic policies. While household surveys can serve as a rich source of socio-economic data, conducting them often entails a great deal of administrative, technical and financial resources. With limited resources for data collection, this often puts pressure on national statistical systems to meet the continuously growing data demand of its stakeholders, especially in developing countries. Using a master sample design that can be used to select samples for multiple household surveys provides an opportunity to minimize the resources needed to collect household survey data regularly. In particular, using the same sampling design and frame to select samples either for multiple surveys of different content or for different rounds of the same survey could induce significant cost-savings instead of developing an independent design each time a household survey is to be carried out. This paper provides a step-by-step guide for developing a master sample design for household surveys in developing countries. Using Bangladesh as a case study, issues like effective sample allocation to ensure the reliability of domain estimates, stratification measures to reduce design effects and introducing household sample size adjustment when to maintain uniform selection probability within domain are discussed.

1. Background

Household surveys have been important source of various socio-economic information that are indispensable in development planning and policy analysis. In some countries, especially developing ones, household surveys have become more dominant form of data collection than other administrative data collection programs such as civil registration systems (UNSD 2005). Thus, there is a need to ensure that household surveys follow scientifically-sound design to assure the quality of information that can be derived from it. While many countries have put in place national statistical systems for collecting household surveys, they have varying levels of experience and infrastructure in data collection. Many developing countries usually confront budgetary constraints and thus, heavily rely on technical assistance from international development agencies. To promote sustainability of statistical data collection activities, different strategies have been proposed to economize the technical and financial resources needed for collecting household surveys. One of these strategies is the development of a master sample design. For multi-stage household surveys, a master sample design allows one or more stages to be combined or shared among different household surveys. In turn, a master sample refers to the sample resulting from the shared stages (UNSD 2005). The UNSD (2005) identified several advantages of adopting a master sample design. First, it reduces costs of developing and maintaining sampling frames as more household surveys share the same master sample design. It also simplifies the technical process of drawing individual samples by facilitating operational linkages between different surveys. In this study, we document our experiences in helping Bangladesh develop a master sample design for collecting household surveys.

The Bangladesh Bureau of Statistics (BBS) is the government agency mandated to undertake data collection for the compilation of official statistics for Bangladesh. Household surveys of national coverage are the primary data collection tool of BBS. Prior to the 2009-2010 Labour Force Survey, the 2005 LFS and 2005 Household Income and Expenditure Survey (HIES) were the last household surveys conducted by BBS, both of which employed the Integrated Multi-Purpose Sampling Design (IMPS). However, previous studies such as that of Maligalig and Barcenas (2008) identified technical deficiencies in IMPS. In particular, large design effects were derived for important characteristics of interest such as unemployment rate in the statistical metropolitan areas (SMA) and for large divisions such as Dhaka, Chittagong and Rajshahi because of ineffective stratification measures. In addition, the survey weights used in IMPS did not reflect the selection probabilities that were applied at the time when the sampled households were drawn. Moreover, Maligalig and Barcenas (2008) also noted that the number of households sampled per primary sampling unit (PSU) can still be reduced and the number of PSUs increased to mitigate the very large design effects. Due to these issues, BBS requested technical assistance from Asian Development Bank in 2008 to develop a new sampling design and sampling frame that could be used for the then forthcoming 2009-2010 Labour Force Survey. Moreover, the main objective of the activity is to re-implement the proposed sampling design once the data from the 2011 Census of Population becomes available. In turn, the updated version will serve as master sample design for the succeeding household surveys that BBS will conduct. This paper documents the processes that were undertaken to develop the sample design for 2009-2010 LFS which will also be the basis for a new master sample design. The study aims to provide survey statisticians from developing countries with empirical guidelines in developing a master survey design for multiple household surveys.

After this introduction, the succeeding discussions are outlined as follows. Section 2 provides a guide for constructing primary sampling units for household surveys. Section 3 identifies the statistical and practical issues that must be considered in designating the survey domain. Section 4 discusses survey stratification as a tool for improving precision of survey estimators. Section 5 discusses sample selection schemes that control for design effects of complex surveys. The last section provides a brief summary of the discussions.

2. Sampling Frame of the Primary Sampling Units

Multi-stage sampling is usually the most appropriate, cost effective and commonly used design for household surveys of national coverage in developing countries. Households (or housing dwellings) are the ultimate sampling units while the primary sampling units (PSUs) are usually clusters of contiguous households. Although stratified simple random sampling is perhaps the most efficient among the conventional sampling designs, it is not practical and workable for most household surveys in developing countries because an updated list of all households in a country is not commonly available. In general, a good sampling frame is needed to ensure that each ultimate sampling unit has a chance of being selected and hence, conclusions on the target population can be drawn from the sample.

Constructing a frame of the primary sampling units is the first step in developing a multi-stage sampling design. At this point, it is important to decide carefully on what should be designated as the PSU. There are several considerations. Ideally, all units in the target population should belong to one and only one of the PSUs. To this end, PSUs must have clear boundaries which can be easily located in the field. In addition, auxiliary information about the "size" of the PSU to be used for selecting which unit will be in the sample should be available. If the total number of households is used as measure of PSU's size, a PSU has to be as manageably small as possible but large enough to have adequate number of ultimate sampling units. This would permit sampling rotations for different surveys which will be implementing the master sample design. Moreover, availability of information to be used for stratification and sample allocation should also be among the practical considerations in constructing the PSUs.

In Bangladesh, unions, mauza, villages and enumeration areas as defined in the 2001 Census of Population are possible candidates for designating PSU. However, preliminary analysis shows that unions vary widely in size and in general, they are too large to permit manageable field operations. Villages, on the other hand, are almost the same as enumeration areas for rural areas and their boundaries are not clear in the case of urban areas. Given this information, only mauza and enumeration areas were considered for PSUs in the succeeding discussions. Using data from the 2001 Census of Population, Tables 1 and 2 summarize the distribution of the number of households by mauza, and by enumeration area, respectively. Noticeably, the total number of households vary widely at the mauza level, from 1 to 22,366 (Other Urban Areas, Dhaka). If mauza is designated as the PSU, then some mauzas will still have to be divided further to ensure that each PSU will not be selected more than once. At the same time, some mauzas may need to be combined to ensure that there is sufficient number of households that can be drawn from each PSU. In contrast, although there are still many enumeration areas that need to be combined if they are designated as PSUs, enumeration areas need not be broken down further since the maximum total number of households per enumeration areas is 497 (Table 2). Thus, forming PSUs using the enumeration areas presents a better option than designating mauzas as PSUs.

Table 1. Summary statistics of total household of Bangladesh by Mauza

Division	No. of Mauzas	Urban block	No. of Mauzas by block	Distribution of the Number of Households by Mauza					
				Total	Min	Median	Mean	Max	Std dev
Barisal	3414	Rural	2,896	1,411,766	1	321	487.49	5,126	481.19
		Urban block	419	144,911	11	249	345.85	3,196	333.86
		SMA	99	91,408	26	718	923.31	4,342	794.92
		Other urban areas	—	—	—	—	—	—	—

Chittagong	8879	Rural	7,367	3,317,141	1	258	450.27	11,943	600.50
		Urban block	1,175	743,076	1	284	632.41	9,831	1,130.94
		SMA	207	257,432	9	794	1,243.63	8,045	1,398.13
		Other urban areas	130	154,899	10	823	1,191.53	5,328	1,208.63
Dhaka	18295	Rural	14,660	5,399,312	1	219	368.30	8,820	473.40
		Urban block	2,616	1,824,745	1	303	697.53	9,218	1,175.43
		SMA	289	254,248	1	576	879.75	5,325	952.55
		Other urban areas	730	758,382	1	316	1,038.88	22,366	2,283.58
Khulna	7483	Rural	6,300	2,472,098	1	264	392.04	5,119	422.51
		Urban block	913	433,156	1	307	474.43	5,823	544.95
		SMA	166	131,468	51	583	791.98	4,101	705.05
		Other urban areas	105	82,880	1	408	789.33	4,938	1,015.84
Rajshahi	18887	Rural	16,423	5,643,537	1	221	343.64	5,758	382.17
		Urban block	1,951	645,620	1	232	330.92	2,597	304.51
		SMA	340	280,392	5	588	824.68	4,042	703.76
		Other urban areas	173	58,248	1	278	336.69	2,026	301.33
Sylhet	5708	Rural	4,989	1,213,085	1	167	243.15	3,052	256.24
		Urban block	608	110,982	1	136	182.54	1,328	166.86
		SMA	111	64,155	20	427	577.97	2,865	559.31
		Other urban areas	—	—	—	—	—	—	—

Source: Authors' computations using data from 2001 Census of Population conducted by BBS

Table 2. Summary statistics of total household of Bangladesh by Mauza

Division	Urban block	No. of enumeration areas	Households					
			Total	Min	Median	Mean	Max	Std dev

Barisal	Rural	14,473	1,411,766	1	96	97.54	354	25.58
	Urban block	1,573	144,911	1	88	92.12	233	29.94
	SMA	898	91,408	2	98	101.79	267	28.13
	Other urban areas	—	—	—	—	—	—	—
Chittagong	Rural	36,172	3,317,141	1	94	91.70	321	34.25
	Urban block	7,943	743,076	1	92	93.55	339	31.73
	SMA	2,997	257,432	1	87	85.90	237	39.46
	Other urban areas	1,428	154,899	2	107	108.47	310	35.52
Dhaka	Rural	54,822	5,399,312	1	99	98.49	483	31.38
	Urban block	18,819	1,824,745	1	93	96.96	471	37.97
	SMA	2,418	254,248	1	102	105.15	404	36.49
	Other urban areas	7,030	758,382	1	100	107.88	478	43.49
Khulna	Rural	23,530	2,472,098	1	104	105.06	320	30.28
	Urban block	3,998	433,156	1	103	108.34	344	35.53
	SMA	1,187	131,468	8	108	110.76	239	31.07
	Other urban areas	744	82,880	1	106	111.40	305	34.66
Rajshahi	Rural	55,004	5,643,537	1	101	102.60	463	29.94
	Urban block	6,707	645,620	1	93	96.26	497	35.07
	SMA	2,639	280,392	1	103	106.25	286	33.12
	Other urban areas	546	58,248	1	104	106.68	295	34.59
Sylhet	Rural	14,875	1,213,085	1	84	81.55	258	36.29
	Urban block	1,302	110,982	1	84.5	85.24	276	39.11
	SMA	723	64,155	1	90	88.73	258	37.99
	Other urban areas	—	—	—	—	—	—	—

As mentioned earlier, it is ideal for every PSU to be large enough to have adequate number of ultimate sampling units to ensure the feasibility of adopting a rotating sample design for different surveys which will be implementing the master sample. In the case of Bangladesh, we set the threshold to be 40 households per PSU. Out of the 259,828 EAs, 12,273 EAs have less than 40 households. These small EAs should be considered as candidates for merging. When combining small EAs to form PSUs, the main consideration is that the enumeration areas to be combined are contiguous. However, due to the lack of reliable (geographic) maps for these EAs, we decided to combine the small enumeration areas based on the criteria provided below. In addition, due to the conceptual and logistical problems in the classification of statistical metropolitan areas (SMA) and other urban areas, it was decided that these two areas will be classified under urban area instead.

Criteria for combining enumeration areas to form a primary sampling unit

1. An EA with more than 40 households is directly considered as a PSU.
2. A small EA is attached to an adjacent EA that belongs to the same urban/rural classification and mauza.

3. A small single EA in a mauza can be combined with an EA of another mauza provided that both mauzas belong to the same union and the EAs to be combined belong to the same urban/rural category.

Following this criteria, a total of 248,904 PSUs were constructed out of the 259,828 original EAs. Table 3 provides the distribution of the number of households by PSU. As shown in this table, there are still PSUs that have less than 40 households. These correspond to cases where the unions were very small in terms of number of households. Since these very small units constitute only of 11 PSUs, we decided to exclude them from the sampling frame.

Table 3. Summary statistics of total household of Bangladesh by PSU

Division	Urban block	No. of PSUs	Households					
			Total	Min	Median	Mean	Max	Std dev
Barisal	Rural	14,280	1,411,766	41	97	98.86	354	24.30
Barisal	Urban	2,414	236,319	42	94	97.90	267	27.36
Chittagong	Rural	33,721	3,317,141	41	97	98.37	321	28.17
Chittagong*	Urban	11,810	1,155,407	23	95	97.84	339	30.98
Dhaka	Rural	52,667	5,399,312	19	100	102.52	483	27.88
Dhaka*	Urban	27,317	2,837,375	21	98	103.88	478	36.93
Khulna	Rural	22,886	2,472,098	31	105	108.02	320	27.01
Khulna	Urban	5,823	647,504	41	105	111.20	344	32.69
Rajshahi	Rural	53,554	5,643,537	41	102	105.38	463	27.28
Rajshahi	Urban	9,614	984,260	13	98	102.38	497	32.20
Sylhet	Rural	12,992	1,213,085	41	92	93.37	266	29.79
Sylhet	Urban	1,826	175,137	21	93	95.91	296	33.27

Source: Authors' computations using data from 2001 Census of Population conducted by BBS.

*Notes: * – There are 3 PSUs that have very few number of identified households on the basis of the latest census data. In particular there are one PSU from Chittagong (urban) and two from Dhaka (urban) that have less than 10 households. These were not included in the computation of summary statistics provided above.*

3. Survey Strata, Determination of Sample size and Sample Allocation

Design domains or explicit strata are subpopulations for which separate samples are planned, designed and selected (Kish, 1987). The choice of explicit strata depends on several factors such as reporting requirements, sampling design and more importantly, available budget and workload that will be used (Kish, 1965;1987). Both statistical and practical issues must be considered in designating the strata. In general, there is now greater demand for statistics at finer levels of disaggregation (Elbers, Lanjouw and Lanjouw 2003). In turn, this would require increasing the number of strata. Since the total sample size is usually determined at the stratum level, increasing the number of strata would necessarily entail increasing the total sample size. Because the workable sampling designs would all involve cluster sampling, the expected design effects should also be considered and used to determine the final sample size. Average design effects for cluster samples is expected to be three or more and hence, the final sample size would have to be increased by this value. However, these things should be contextualized within the available budget allocated for survey data collection.

Once the strata have been clearly specified, the sample size for each stratum is then determined so that reliable estimates at the stratum level can be derived. Information on the variability of the sampling units within each stratum, the acceptable error level, and the associated costs are the factors needed to determine the sample size. For example, suppose the primary characteristic of interest to be measured can be expressed as a proportion. Under simple random sampling (SRS), the tentative sample size for a particular stratum is computed such that

(1) $$n_{srs} = \frac{\frac{t^2_{(\alpha,N-1)}P(1-P)}{d^2}}{1+\frac{1}{N}\left(\frac{t^2_{(\alpha,N-1)}P(1-P)}{d^2}-1\right)},$$

where $t_{(\alpha,N-1)}$ is the abscissa of the t-distribution given risk α, and the population size N; P is the true proportion of the characteristic of interest and error level d (Cochran, 1977). Since P is unknown, we can either set $P=0.5$ or used any prior information about the value of P from previous studies. Note that setting $P=0.5$ would produce the most conservative or largest sample size. The resulting sample sizes on the application of (1) are then inflated by the corresponding design effects (Deff), assuming that prior information about the magnitude of the design effect is available.

(2) $$n_{complex} = \text{Deff} * n_{SRS}.$$

In the case of Bangladesh, the geographic divisions were designated as the design domains or explicit strata. If the 64 zilas (provinces) were specified as the strata, the sample size required will be inflated by approximately ten times and that would be beyond the budget of BBS. Moreover, we used the estimated unemployment rate using the 2005 LFS to provide a value for P. Table 4 shows the tentative sample sizes that were computed at risk $\alpha=0.05$ and varying error level d. The corresponding design effects of unemployment rates from the 2005 LFS are also shown in Table 4. Note that at $d=0.01$, the total sample size is 115,277. This sample size is 100,000 more households than what the budget of BBS has allocated for the 2009-2010 LFS can afford. At $d=0.03$, which may not be very appropriate considering that unemployment rates are quite small, the total sample size is about 12,814 households or within budget. The total sample size became very large because of large design effects especially for Dhaka and Khulna. The perceived large variability among these divisions may not really reflect the large variation across households in these divisions but the wide variation in the artificial weights that were attributed to the households. Given this backdrop, the sample sizes that were computed in Table 4 were used only as guides for determining the final total sample size. In particular, we proposed to sample 10 households per PSU following the recommendation of Maligalig and Barcenas (2008) instead of the 40 households per PSU followed in IMPS. This allows us to increase the number of sampled PSU from 1000 in IMPS to 1500 in the new sample design. Considering that there is positive intra-correlation among households in the same PSU, then increasing the number of sampled PSU while reducing the number of sampled household per PSU is deemed reasonable.

If the survey weights used to compute the sample size in Table 4 were correct, the estimates at the division level will have margin of error of about .03. This is not acceptable since this error level is quite large considering that division level unemployment rates only varies from .01 (Sylhet) to .06 (Barisal). On the other hand, since the survey weights in the 2005 LFS have technical flaws and stratification measures used were not effective in controlling the design effects, the resulting estimates from the 2009-2010 LFS using the proposed master sample design can still render acceptable design effects even with only 15,000 households total sample size. This favorable outcome depends on the quality of implementation of a better design for the master sample, specification of the correct survey weights and better stratification.

Table 4. Tentative Sample Sizes

Division	Unemployment Rate	# of households	DEFF	SRS sample size			Sample size: Complex survey		
				d=0.05	d=0.03	d=0.01	d=0.05	d=0.03	d=0.01
Barisal	0.0622	1,648,085	5.12	89.57	248.77	2236.24	460.31	1278.51	11492.75
Chittagong	0.0461	4,472,548	8.38	67.51	187.53	1687.17	567.31	1575.81	14177.52
Dhaka	0.0474	8,236,687	27.00	69.37	192.70	1733.99	1878.49	5217.95	46952.74
Khuha	0.0545	3,119,602	18.58	79.18	219.92	1978.19	1475.64	4098.83	36868.64
Rajshahi	0.0311	6,627,797	3.41	46.26	128.51	1156.41	158.07	439.08	3951.07
Sylhet	0.0182	1,388,222	2.66	27.53	76.47	687.90	73.40	203.88	1834.07
Total							4613.22	12814.05	115276.79

Source: Authors' computations using data from 2005 LFS conducted by BBS.

Several allocation strategies were examined to allocate the 15,000 sample households across domains: equal allocation, proportional allocation, square root allocation and Kish allocation.

Equal Allocation:

$$n_d = \frac{n}{D} = \frac{n}{6}$$

Proportional Allocation:

$$n_d = n\frac{N_d}{N} = nW_d$$

Square Root Allocation:

$$n_d = n\frac{\sqrt{N_d}}{\sum_m \sqrt{N_m}}$$

Kish Allocation:

$$n_d = n \frac{\sqrt{D^{-2} + IW_d^2}}{\sum_m \sqrt{D^{-2} + IW_m^2}} = n \frac{\sqrt{\frac{1}{36} + IW_d^2}}{\sum_m \sqrt{\frac{1}{36} + IW_m^2}}$$

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 Bangladesh, per the 2001 Census of Population, 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 (type (i)) as compared to those estimates at the domain levels (type (ii)). To illustrate, we can relate (i) to characteristics of interest such as numbers of crop farmers and female unpaid workers, proportions of persons in poverty in Bangladesh, number of persons in the labor force who are unemployed, proportion of households with electricity, and estimates of the differences between subgroups. When computed at the domain level, these become type (ii) parameters. If the primary interest is to derive estimates for characteristics of interest of type (ii), one of the best approaches in allocating the total sample size is to allocate it proportionally with respect to the population size of each domain. However, the ideal approach for type (ii) is to divide the total sample size equally among the domains (Kish, 1987). Moreover, it should be emphasized that these two approaches may yield very different sample allocations particularly when the domains differ in measure of size. Further, it is possible that a particular approach may perform satisfactorily when estimating a certain type of characteristic of interest but not necessarily for the other types. A possible way around this problem is to use Kish allocation which is basically a compromise between equal and proportional allocation. With $I = 0$, it reduces to the equal allocation while it tends to proportional allocation approach with $I \to \infty$. Table 5 provides estimates of sample size per domain using different allocation procedures. Kish allocation at $I = 1$ was chosen to ensure that precision of both type (i) and type (ii) characteristics of interest will be approximately the same.

Table 5. Sample Allocation of Number of Sample Households per Domain

Division (d)	Total Households N_d	W_d	Equal Allocation $\frac{n}{6}$	Proportional Allocation nW_d	Square root Allocation $n\frac{\sqrt{N_d}}{\sum_m \sqrt{N_m}}$	Kish Allocation ($I = 1$) $n\frac{\sqrt{\frac{1}{36}+W_d^2}}{\sum_m \sqrt{\frac{1}{36}+W_m^2}}$
Barisal	1,648,085	0.064649	2,500	969.73	1,633.65	1,817.68
Chittagong	4,472,548	0.175443	2,500	2,631.64	2,691.21	2,460.51
Dhaka	8,236,687	0.323097	2,500	4,846.45	3,652.13	3,696.56
Khulna	3,119,602	0.122371	2,500	1,835.57	2,247.60	2,102.39
Rajshahi	6,627,797	0.259986	2,500	3,899.78	3,276.08	3,140.06
Sylhet	1,388,222	0.054455	2,500	816.83	1,499.34	1,782.81
Bangladesh	25,492,941	1.000000	15,000	15,000.00	15,000.00	15,000.00

Source: Authors' computations using different sample allocation procedure.

4. Implicit stratification of Primary Sampling Units

(Implicit) Stratification of PSUs is critical to ensuring that the (limited) sample size afforded by BBS will still render reliable estimates at the domain level and those that cut across domains. Ideally, a implicit stratification measure should be available and measured consistently for all PSUs in the domain. Examples of such stratification measures are geographical information such as zila (provinces) and urban/rural areas since each PSU carry the provincial code as well as the urban/area classification. Further stratification may be applied to ensure that the final groups of PSUs are more homogeneous. The candidates for stratification measures that are available for all PSUs are those variables that are in the

2001 Census of Population. In addition, an effective stratification measure is one that is highly correlated with major characteristics of interest in the survey. Those perceived to be correlated to income and employment which are the major characteristics of interests in LFS includes the proportion of households with strong housing materials (PStrong), proportion of households with agriculture as major source of income (PAgri); and proportion of households that own agricultural land (POal). Table 6 present the summary statistics for these three variables by division and rural/urban classification.

Table 6. Summary Statistics of Stratification Measures by Division and Urban/Rural

Division	Stratification Measures	Urban/Rural	Minimum	Median	Mean	Max	Standard Deviation
Barisal	PStrong	Rural	0	0.99	2.93	100	7.11
		Urban	0	14.93	25.37	100	26.45
	PAgri	Rural	0	61.68	59.75	100	23.59
		Urban	0	7.75	20.33	100	24.91
	POal	Rural	0	69.46	66.26	100	22.92
		Urban	0	50.57	50.46	100	23.62
Chittagong	PStrong	Rural	0	4.05	7.46	100	10.70
		Urban	0	30.48	38.11	100	31.68
	PAgri	Rural	0	46.94	48.99	100	25.85
		Urban	0	4.55	15.21	100	22.29
	POal	Rural	0	58.33	57.26	100	22.87
		Urban	0	38.63	40.41	100	25.21
Dhaka	PStrong	Rural	0	1.85	5.37	100	9.49
		Urban	0	57.56	53.90	100	35.59
	PAgri	Rural	0	67.42	62.93	100	24.50
		Urban	0	1.25	10.24	100	19.36
	POal	Rural	0	61.54	61.37	100	20.77
		Urban	0	48.54	48.32	100	26.22
Khulna	PStrong	Rural	0	15.27	17.87	100	14.37
		Urban	0	44.17	46.28	100	27.19
	PAgri	Rural	0	71.07	65.90	100	22.95
		Urban	0	6.49	19.61	100	25.59
	POal	Rural	0	61.17	60.87	100	20.36
		Urban	0	43.33	44.54	100	22.60

Rajshahi	PStrong	Rural	0	3.80	7.68	100	11.05
		Urban	0	33.33	39.39	100	30.84
	PAgri	Rural	0	76.09	70.46	100	22.39
		Urban	0	12.00	24.86	100	27.29
	POal	Rural	0	57.03	57.14	100	19.51
		Urban	0	39.39	40.72	100	20.87
Sylhet	PStrong	Rural	0	11.83	17.92	100	18.68
		Urban	0	49.07	47.42	100	29.50
	PAgri	Rural	0	58.76	56.09	100	28.57
		Urban	0	7.25	18.55	100	23.43
	POal	Rural	0	49.38	49.36	100	22.63
		Urban	0	38.65	40.68	100	22.71
Bangladesh	PStrong	All	0	6.06	17.47	100	25.36
	PAgri	All	0	56.82	51.12	100	31.86
	POal	All	0	56.43	55.64	100	23.06

Source: Authors' computations using data from 2001 Census of Population conducted by BBS.

There are several findings that may be indicative that the urban/rural classification should be reviewed carefully. In particular, there are PSUs for urban areas in which all households have agriculture as main source of income while there are PSUs in rural areas with not even one household that has agriculture as main source of income. Table 6 also shows that ownership of agricultural land is not a very good distinguishing factor for urban/rural areas. This probably shows that there are many owners in urban areas who rent or lease their agricultural land and hence, decreasing the value of POal as a stratification measure.

As indicated by the standard deviation, minimum, median and maximum values, PStrong does not vary widely in rural areas. On the average, there is considerably much lower proportion of households that have strong housing materials in the rural areas. On the other hand, although the variation of PAgri is about the same for urban and rural areas in some divisions, the number of households with agriculture as main source of income is significantly much lower in the urban areas, on the average. These results prompted us to stratify urban areas using PStrong and rural areas using PAgri. In particular, since the numbers of households and PSUs in rural areas are more than twice those of the urban areas, four and two strata were planned for rural and urban areas, respectively. Strata boundaries were first set as the quartiles of PAgri for rural areas and the median of PStrong for urban areas. However, small strata or those that have total households that is less than the division's sampling interval are combined with the adjacent strata. The number of PSUs for each of the 336 strata that were formed are summarized in Appendix 3.

In general, the key advantage of the (implicit) stratification procedure adopted here is that it is straightforward to implement and provide satisfactory results. Nevertheless, future studies may consider implementing more optimal stratification procedures such as those proposed by Sethi (1963) and Kozak (2004).

5. Sample Selection

Another measure for controlling design effect is to ensure that the survey weights within the domains do not vary widely. A wide variation of weights within a domain will unnecessarily contribute to the increase of variances of estimates. Hence, survey statisticians usually opt to maintain almost similar base weights within a domain. Since base weight is the inverse of the selection probability of an ultimate sampling unit, then maintaining similar or almost uniform base weights is tantamount to maintaining the same or almost the same selection probabilities within a domain. This section discusses the procedures on how this can be achieved. Here, we propose a simple two-stage sampling design such that in a domain d: (i) PSU α will be selected with probability proportional to size and (ii) household β from PSU α will be selected by simple random or systematic sampling, in a domain d in which all PSUs are also grouped into implicit strata. Thus, in domain d and (implicit) stratum h, the uniform selection probability f_d that a household is selected from PSU α will be:

(3) $$f_d = \frac{n'_d}{M_d},$$

,

where n'_d is the total sample size for domain d as defined in the last column of Table 5 (Kish Allocation, Index=1), M_d is the measure of size for domain d (i.e., total number of households per division based on the 2001 Census of Population data) and $M_{h\alpha}$ is the measure of size for PSU α at stratum h (i.e., total number of households for PSU α from stratum h),

(4) $$M_d = \sum_h \sum_\alpha M_{h\alpha}.$$

In a two-stage cluster sampling design,

(5) $$f_d = P(\alpha\beta|h) = P(\alpha|h)P(\beta|h\alpha),$$

where $P(\alpha|)$ is the selection probability of PSU α and $P(\beta|h\alpha)$ is the probability of selecting household β given PSU α in stratum h is selected. Hence,

(6) $$f_d = \frac{a_h M_{h\alpha}}{\sum M_{h\alpha}} \frac{b_h}{M_{h\alpha}} = \frac{a_h b_h}{\sum M_{h\alpha}},$$

where a_h is the number of PSUs to be sampled from stratum h, and b_h is the number of households to be selected from stratum h.

The term $P(h\beta|\alpha) = \frac{b_h}{M_{h\alpha}}$ represents the sampling fraction to be used in the systematic sampling of households at the final sampling stage. Its inverse is the sampling interval to be applied in the selection of households from the sampled PSUs.

Considering (6), f_d will be uniform in a domain when $\frac{a_h}{\sum M_{h\alpha}}$ and b_h do not depend on stratum h and hence, are both constant across all strata in domain d. Since the recommendation that $b_h = 10$ for all sampled PSUs will be implemented, and if $\frac{a_h}{\sum M_{h\alpha}}$ can be maintained to remain constant, f_d will be uniform in domain d. To do latter, the number of PSUs to be selected for stratum h, a_h must be proportional to the stratum h measure of size $\sum M_{h\alpha}$, which is actually the 2001 Census of Population total number of households for stratum h. However, since a_h must be a whole number and the strata measure of sizes also vary, the resulting selection probabilities across strata in domain d will not be totally the same but will not vary widely.

To maintain a uniform f_d in the whole domain, the same sampling interval can be applied on the list of all PSUs that are already sorted by strata. This implies that the selection of PSUs will not be done separately for each stratum in a domain but rather, will be performed collectively for all of the strata. The step-by-step procedure for maintaining a uniform selection probability within the domain is outlined below. Table 7 below shows the resulting uniform selection probabilities for each domain.

Sample Selection of Primary Sampling Units

(1) For a domain d, determine the number of PSUs to be sampled $a_d^{'}$, such that $a_d^{'} = \frac{n_d^{'}}{b}$, where b is the recommended number of households per PSU (in this case, b=10), $n_d^{'}$ is the number of households allocated to domain d (Table 5, last column).

(2) Compute the sampling interval:

(7) $$\frac{\sum_\alpha M_{h\alpha}}{a_d^{'}} = \frac{N_d}{a_d^{'}}.$$

(3) Sort all the PSUs in domain d by zila, urban/rural classification, by strata and lastly, by PStrong values.

(4) Compute the cumulative value of the measure of size (total number of households based from 2001 Census of Population), $M_{h\alpha}$ using the sorted list in step (3).

(5) Select a random start (RS) by drawing a random number between 0 and 1 and multiplying it by the interval in step 2. The first sampled PSU will be the first PSU with cumulative value of $M_{h\alpha}$ containing the value of the random start (RS). The next sample PSU will be the PSU for which the cumulative value of $M_{h\alpha}$ contains $RS + S_d$, the next will be the PSU for which the cumulative value contains $RS + 2 * S_d$, etc.

Table 7. Summary of Sample Statistics by Domain

Division	Total No. of Households N_d	Computed Sample PSUs $a_d^{'}$	Sampling Interval S_d	Actual Number of sample PSUs a_d	Tentative Sample Households $hatn_d$	Selection Probability f_d
Barisal	1,648,085	181.77	9066.992	182	1820	0.001104

Chittagong	4,472,548	246.05	18177.35	246	2460	0.000550
Dhaka	8,236,687	369.66	22282.06	370	3700	0.000449
Khulna	3,119,602	210.24	14838.39	210	2100	0.000673
Rajshahi	6,627,797	314.01	21107.21	314	3140	0.000474
Sylhet	1,388,222	178.28	7786.691	178	1780	0.001282

Source: Authors' computations using data from 2001 Census of Population conducted by BBS.

Sample Selection of Households

Since the measure of size (i.e., total number of households) that was used for selecting the PSUs is based on 2001 Census of Population which is quite far from the 2009-2010 reference period of the LFS, the number of households to be sampled must be adjusted accordingly to maintain the uniform selection probabilities within domain. In particular, since the households will be selected from a sampled PSU α with $P(h\beta|\alpha) = \frac{b_h}{M_{h\alpha}}$ and if the 2009-2010 value of the measure of size is denoted as $N_{h\alpha}$, then maintaining the same household level selection probability means that

(8) $$P(h\beta|\alpha) = \frac{b_h}{M_{h\alpha}} = \frac{10}{M_{h\alpha}} = \frac{b'_{h\alpha}}{N_{h\alpha}},$$

and hence,

(9) $$b'_{h\alpha} = \frac{N_{h\alpha}}{M_{h\alpha}} * 10,$$

where $b'_{h\alpha}$ is the actual total number of households to be selected in PSU α in stratum h. This implies that the there should be a listing operation of all households in the selected PSUs before the conduct of the 2009-2010 LFS.

6. Survey Weights and Estimation

The complex design of the master sample has to be considered in analyzing the 2009-2010 LFS and other surveys that will use the master sample in the future. Survey weights must be used to produce estimates of population parameters and design features such as the stratification measures, PSUs and domains must be taken into account in variance estimation and inference.

6.1 Survey Weights

The final survey weights are the product of at most three successive stages of computations. First, base weights are computed to counteract the unequal selection probabilities in the sample design. Then the base weights are adjusted to balance uneven response rates and if data are available, the non-response adjusted weights are further adjusted to ensure that the weighted sample distributions conform with known distributions from valid auxiliary data sources.

The base weight for sampled household is the inverse of its selection probability. In the master sample design, the selection probability is uniform within a domain and hence, base weights will not also vary within domains. In general,

(10) $$w_d^0 = \frac{1}{f_d}.$$

Table 8 presents the base weights of sampled households by division.

Table 8. Base Weights by Domain

Division	Selection Probability f_d	Base Weight w_d^0
Barisal	0.001104	905.7971
Chittagong	0.000550	1818.1820
Dhaka	0.000449	2227.1710
Khulna	0.000673	1485.8840
Rajshahi	0.000474	2109.7050
Sylhet	0.001282	780.0312

Source: Authors' computations using data from 2001 Census of Population conducted by BBS

Non-response adjustments will have to be incorporated in the final survey weights if the degree of unit non-response cannot be ignored. Unit non-response occurs when an eligible household fails to participate in the survey. For example, households may refuse to participate or an eligible respondent may not be available at the times that the survey interviewer visits. In general, the non-response adjustment inflates the base weights of "similar" responding units to compensate for the non-respondents. The most common form of non-response weighting adjustment is a weighting class type. The full sample of respondents and non-respondents is divided into a number of weighting classes or cells and non-response adjustment factors are computed for each cell c (Kalton, 1990) as

(11) $$w_c^1 = \frac{\sum_{i\epsilon rc} w_{di} + \sum_{j\epsilon mc} w_{dj}}{\sum_{i\epsilon rc} w_{di}} = \frac{\sum_{i\epsilon sc} w_{di}}{\sum_{i\epsilon rc} w_{di}},$$

where the denominator of w_c^1 is the sum of the weights of respondents (indexed r) in weighting cell c while the numerator adds together the sum of the weights for respondents and the sum of the weights for eligible non-respondents (indexed m for missing) in cell c which is equal to the sum of the weights for the total eligible sample (indexed s) in cell c. Thus, the non-response weight adjustment w_c^1 is the inverse of the weighted response rate in cell c. Note that the adjustment is applied with eligible units. Ineligible sampled units (e.g., vacant or demolished housing units and units out of scope for a given survey) are excluded.

Weighting cells c need not conform with the strata boundaries. They may cut across strata but it is important that the weighting cells will capture "similar" households. Similarity is viewed here in the perspective of the households propensity to response. In general, the response rates across weighting cells will vary widely. Moreover, there may be instances that the weighted sample distributions will not conform with projected population counts. When this happens, further weighting adjustments or what is known as population weighting adjustments can be incorporated in the final survey weight to ensure that the sample distribution conforms with the population distribution. Population weighting adjustment is performed similar to the non-response weighting adjustments described earlier. Calibration methods such as raking are used in this process. Using an iterative proportional fitting algorithm, raking is performed on the non-response adjusted weights such that the weighted survey estimates of some characteristics of interest (e.g. age group and sex) conform with the corresponding population distributions.

6.2 Estimation

Assuming that the final survey weight for household i is w_i or what can be viewed as the number of population units that the responding household i represent. Then the estimator of a population total for characteristic of interest Y will be $\hat{Y} = \sum_{i\epsilon s} w_i y_i$, where y_i is the value of the variable y for household i.

The simple estimator $\hat{Y}$ has many applications. For example, it can be applied to estimate the count of population with specific characteristic of interest, by setting $y_i = 1$ if household i has the specific characteristic, 0 otherwise.

To estimate the population mean, $\bar{Y}$, the following ratio estimator can be used:

(12) $$\bar{y} = \frac{\sum_{i\epsilon s} w_i y_i}{\sum_{i\epsilon s} w_i},$$

with the total of the survey weights of all responding households, $\sum_{i\in s} w_i$, as an estimator for the total number of households. A more general form of the ratio estimator (Kalton, 1983) would be:

(13) $$R = \frac{\sum_{i\epsilon s} w_i y_i}{\sum_{i\epsilon s} w_i x_i}.$$

Note that with complex sample design such as the master sample, the means depicted in (12) and (13) are ratio estimators that involve the ratio of two random variables and hence, must be carefully considered in the computation of sampling errors.

6.3 Variance Estimation

The variances of survey estimates are needed to evaluate the precision of the survey. The sampling design in addition to the sample size is critical to the precision of survey estimates. The statistical software packages have modules that can approximate the variance of estimates from complex surveys. Most of these software packages make use of the Taylor series approach in computing the variance, although some software also offers alternate approach in the form of replication, resampling or bootstrap procedures. In general, each variance estimation approach has its own advantages and limitations. For instance, while Taylor series expansion approach is more straightforward to implement, incorporating non-response adjustments may render this technique less appropriate. In such context, resampling procedures may give more accurate approximations of the true variance. Nevertheless, in all these variance estimation techniques, specifying the features of the survey design is required. Also, these approaches involve approximations, most are anchored on the assumption that the first stage sampling fractions are small.

Note that survey estimates at the (geographic) division level are expected to have sampling error at acceptable level. This is also expected for estimates at the national level that cut across domains. For example, unemployment rates at urban/rural area levels are expected to have tolerable sampling errors. It is important that sampling errors of major estimates should be derived to validate these expectations. Moreover, sampling errors are also needed to evaluate the reliability of estimates at the sub-division level (e.g., zila level in the case of Bangladesh). Estimates for sub-division with sufficient sample size may render acceptable sampling errors. In the case of Bangladesh, some zilas still have relatively large sample size. Thus, although the divisions are set as the design domains or explicit strata, some estimates at the zila level may still have tolerable sampling error. However, disaggregating zila-level estimates by urban/rural may not at all be possible because of insufficient sample size.

7. Summary

The paper documents the technical processes that were undertaken in the development of the new sample design that was used for the 2009-2010 Labor Force Survey conducted in Bangladesh. The new sample design addresses the weaknesses identified in the previous design adopted in 2005 LFS. Some of the (proposed) changes are as follows: first, considering the positive intra class correlations of major characteristics of interest, the total number of households to be enumerated per was reduce from 40 to 10 while the number of PSUs to be selected was increased from 1000 to 1500. Second, effective sample allocation procedure was implemented to ensure the reliability of estimates at the division-level as well as those that cut across divisions. Third, implicit stratification measures were introduced to reduce design effects. Fourth, a sample selection procedure that maintains uniform selection probability for each division was also adopted to counter the large design effects noted from 2005 LFS.

Appendix 1

The Integrated Multi-Purpose Sample Design

The Integrated Multi-Purpose Sample Design (IMPS) was used by the Bangladesh Bureau of Statistics (BBS) to sample households for surveys of national coverage. Two such surveys are the 2005-06 Labour Force Survey (LFS) and the 2005 Household Income and Expenditure Survey. In general, IMPS has a stratified cluster design. Clusters of about 200 households each were formed as enumeration blocks for each zila (municipality) on the basis of the 2001 Census of Population. These enumeration blocks served as the primary sampling units (PSUs) in IMPS and were classified as urban, rural and statistical metropolitan areas (SMA). Further geographical stratification were also introduced by classifying the zilas according to six divisions – Barisabal, Chittagong, Dhaka, Khulna, Rajshahi and Sylhet. In all, there were 129 strata formed – 64 strata corresponding to 64 rural zilas, 61 strata classified under urban with the other three, Gazipur, Narayanganj and Khulna taken together to form one strata under SMA in addition to the other three SMA strata formed from urban areas with very large population – Dhaka, Chittagong, Rajshahi.

Of the 109,000 (?) PSUs, 1000 were selected. The procedure for allocating the PSUs to the 129 strata was not clarified in the documentation. Appendix 1 presents the distribution of the PSUs to the strata. Moreover, the procedure for selecting the PSUs was not also included in the documentation. For each selected PSUs, 40 households were selected at random making the total sample households equal to 40,000.

The survey weight usually derived as the product of the base weight (equal to the inverse of the selection probability) and the adjustments for non-response and non-coverage, was not determined as such. Instead, the survey weight was derived as the ratio of total households in the strata (updated as of April 2006) to the sample households. Appendix 2 presents the survey weights that were derived.
Summary of PSU Allocation Across Strata

Strata	National	Rural	Urban	SMA
Barisal Division	**80**	**55**	**25**	–
06- Barisal zila	17	12	5	–
09- Bhola zila	14	10	4	–
42- Jhalokati zila	12	8	4	–
79- Perojpur zila	12	8	4	–
04- Barguna zila	12	8	4	–
78- Patuakhali zila	13	9	4	–
Chittagong Division	**179**	**116**	**49**	**14**
03- Bandarban zila	12	8	4	–
15- Chittagong zila	34	16	4	14
22- Cox's Bazar zila	12	8	4	–
12- Brahmanbaria zila	15	10	5	–
13- Chandpur zila	15	10	5	–
19- Comilla zila	26	20	6	–
46- Khagrachhari zila	12	8	4	–
30- Feni zila	12	8	4	–
51- Lakshmipur zila	12	8	4	–
75- Noakhali zila	17	12	5	–
84- Rangamati zila	12	8	4	–
Dhaka division	**289**	**172**	**73**	**44**
26- Dhaka zila	34	8	4	22
33- Gazipur zila	18	8	–	10
56- Manikganj zila	12	8	4	–
59- Munshiganj zila	12	8	4	–
67- Narayanganj zila	20	8	–	12
68- Narshingdi zila	15	9	6	–

29- Faridpur zila	14	10	4	–
35- Gopalganj zila	12	8	4	–
54- Madaripur zila	12	8	4	–
82- Rajbari zila	12	8	4	–
86- Shariatpur zila	12	8	4	–
39- Jamalpur zila	15	10	5	–
89- Sherpur zila	13	9	4	–
48- Kishoreganj zila	17	12	5	–
61- Mymensingh zila	33	23	10	–
72- Netrokona zila	14	10	4	–
93- Tangail zila	24	17	7	–
Khulna division	**146**	**89**	**45**	**12**
41- Jessore zila	20	12	8	–
44- Jhenaidah zila	15	9	6	–
55- Magura zila	12	8	4	–
65- Narail zila	12	8	4	–
01- Bagerhat zila	13	8	5	–
47- Khulna zila	20	8	–	12
87- Satkhira zila	14	10	4	–
18- Chuadanga zila	13	8	5	–
50- Kushtia zila	15	10	5	–
57- Meherpur zila	12	8	4	–
Rajshahi division	**251**	**170**	**71**	**10**
10- Bogra zila	21	16	5	–
38- Joypurhat zila	12	8	4	–
27- Dinajpur zila	18	13	5	–
77- Panchagar zila	12	8	4	–
94- Thakurgaon zila	12	8	4	–
76- Pabna zila	16	10	6	–
88- Sirajganj zila	18	13	5	–
64- Naogaon zila	17	13	4	–
69- Natore zila	14	10	4	–
70- Nowabganj zila	12	8	4	–
81- Rajshahi zila	24	10	4	10
32- Gaibandha zila	16	12	4	–
49- Kurigram zila	15	10	5	–
52- Lalmonirhat zila	12	8	4	–
73- Nilphamari zila	13	9	4	–
85- Rangpur zila	19	14	5	–
Sylhet division	**55**	**38**	**17**	**–**

36- Hobiganj zila	13	9	4	–
58- Maulvibazar zila	13	9	4	–
90- Sunamganj zila	14	10	4	–
91- Sylhet zila	15	10	5	–
Total	**1000**	**640**	**280**	**80**

Appendix 2

Integrated Multi Purpose Sampling Design

Survey Weights by Stratum

Stratum		Total updated households			Sample households			Sampling weights		
		Rural	Urban	SMA	Rural	Urban	SMA	Rural	Urban	SMA
06	Barisal	442170	94384	0	480	200		921.19	471.92	0.00
09	Bhola	327262	61052	0	400	160		818.15	381.58	0.00
42	Jhalokati	130186	27988	0	320	160		406.83	174.92	0.00
79	Perojpur	208895	42063	0	320	160		652.80	262.89	0.00
04	Barguna	169695	22626	0	320	160		530.30	141.42	0.00
78	Patuakhali	303056	29761	0	360	160		841.82	186.00	0.00
03	Bandarban	47290	18186	0	320	160		147.78	113.66	0.00
15	Chittagong	749021	19916	840746	640	120	600	1170.34	165.97	1401.24
22	Cox’s Bazar	347072	45420	0	320	160		1084.60	283.87	0.00
12	Brahmanbaria	469347	61421	0	400	200		1173.37	307.11	0.00
13	Chandpur	439039	61017	0	400	200		1097.60	305.08	0.00
19	Comilla	933277	95049	0	800	240		1166.59	396.04	0.00
46	Khagrachhari	80753	31643	0	320	160		252.35	197.77	0.00
30	Feni	235576	33292	0	320	160		736.17	208.08	0.00
51	Lakshmipur	286643	44258	0	320	160		895.76	276.61	0.00

75	Noakhali	520274	55253	0	480	200		1083.91	276.27	0.00
84	Rangamati	79856	34145	0	320	160		249.55	213.41	0.00
26	Dhaka	167198	3491	2191848	320	160	880	522.50	21.82	2490.74
33	Gazipur	257960	0	247896	320	0	400	806.13	0.00	619.74
56	Manikganj	274091	21605	0	320	160		856.53	135.03	0.00
59	Munshiganj	255236	37078	0	320	160		797.61	231.73	0.00
67	Narayanganj	222924	0	331883	320	0	480	696.64	0.00	691.42
68	Narshingdi	350319	80406	0	360	240		973.11	335.03	0.00
29	Faridpur	346816	48608	0	400	160		867.04	303.80	0.00
35	Gopalganj	237833	23708	0	320	160		743.23	148.18	0.00
54	Madaripur	225717	30776	0	320	160		705.36	192.36	0.00
82	Rajbari	190272	25943	0	320	160		594.59	162.15	0.00
86	Shariatpur	223253	22255	0	320	160		697.66	139.09	0.00
39	Jamalpur	397902	79792	0	400	200		994.76	398.96	0.00
89	Sherpur	255789	31996	0	360	160		710.53	199.98	0.00
48	Kishoreganj	505921	74145	0	480	200		1054.00	370.73	0.00
61	Mymensingh	875150	136206	0	920	400		951.25	340.52	0.00
72	Netrokona	406897	40560	0	400	160		1017.25	253.50	0.00
93	Tangail	646284	93884	0	680	280		950.42	335.30	0.00
41	Jessore	470209	99110	0	480	320		979.60	309.72	0.00
44	Jhenaidah	314635	46953	0	360	240		873.98	195.63	0.00
55	Magura	167265	22115	0	320	160		522.70	138.22	0.00
65	Narail	144385	15628	0	320	160		451.20	97.67	0.00
01	Bagerhat	293772	55545	0	320	200		918.04	277.72	0.00

47	khulna	255885	0	327109	320	0	480	799.64	0.00	681.47
87	Satkhira	394005	30802	0	400	160		985.02	192.51	0.00
18	Chuadanga	170231	61525	0	320	200		531.97	307.63	0.00
50	Kushtia	360554	39471	0	400	200		901.39	197.35	0.00
57	Meherpur	120478	14850	0	320	160		376.49	92.82	0.00
10	Bogra	603687	93314	0	640	200		943.26	466.57	0.00
38	Joypurhat	179002	18762	0	320	160		559.38	117.26	0.00
27	Dinajpur	526401	84258	0	520	200		1012.31	421.30	0.00
77	Panchagar	172454	20929	0	320	160		538.92	130.81	0.00
94	Thakurgaon	257353	22835	0	320	160		804.22	142.72	0.00
76	Pabna	389278	112917	0	400	240		973.19	470.49	0.00
88	Sirajganj	549959	67454	0	520	200		1057.61	337.27	0.00
64	Nogaon	503878	46560	0	520	160		969.00	290.99	0.00
69	Natore	300966	49842	0	400	160		752.41	311.51	0.00
70	Nawabganj	252512	77584	0	320	160		789.10	484.90	0.00
81	Rajshahi	347804	26871	166673	400	160	400	869.51	167.94	416.68
32	Gaibandha	448228	43073	0	480	160		933.81	269.21	0.00
49	Kurigram	347381	59961	0	400	200		868.46	299.81	0.00
52	Lalmonirhat	222298	32851	0	320	160		694.68	205.32	0.00
73	Nilphamari	314999	46125	0	360	160		874.99	288.29	0.00
85	Rangpur	489418	95717	0	560	200		873.96	478.59	0.00
36	Habiganj	359402	47883	0	360	160		998.34	299.27	0.00
58	Maulvibazar	339274	34324	0	360	160		942.43	214.53	0.00

90	Sunamganj	413830	48966	0	400	160		1034.58	306.03	0.00
91	Sylhet	482356	114533	0	400	200		1205.90	572.67	0.00

2009 Master Sample

PSU Count by Division, Zila and Urban/Rural Classification

Division	Zila	Rural				Urban		Total
		1	2	3	4	1	2	
Barisal	Barguna	231	386	497	541	214		1869
	Barisal	1328	1046	841	780	314	511	4820
	Bhola	434	604	824	1034	313	170	3379
	Jhaloka	425	392	227	116	105	96	1361
	Patuakh	553	613	662	679	173	175	2855
	Pirojpu	599	525	521	422	184	159	2410
Chittagong	Bandarb	236			421	253		910
	Brahman	498	763	1074	1171	340	231	4077
	Chandpu	738	1160	1084	704	605		4291
	Chittagong	2344	1358	1126	702	2321	4306	12157
	Comilla	1203	2032	2428	1712	518	466	8359
	Cox's B	460	481	698	828	444		2911
	Feni	870	579	445		285		2179
	Khagrac	364			669	452		1485
	Lakshmi	577	693	547	654	424		2895
	Noakhal	1476	1133	734	828	706		4877
	Rangama	311			620	459		1390

Dhaka	Dhaka	756	274	366		5530	10919	17845
	Faridpu	693	762	789	779	450		3473
	Gazipur	698	629	505	384	1099	577	3892
	Gopalga	429	539	474	500	204		2146
	Jamalpu	447	989	1229	1347	761		4773
	Kishorg	988	1252	1170	1281	759		5450
	Madarip	412	515	497	556	270		2250
	Manikga	793	753	620	471	213		2850
	Munshig	1234	520	335		265		2354
	Mymensi	993	1920	2371	2402	938	358	8982
	Narayan	1443	302			1367	899	4011
	Narsing	1372	867	524	313	658		3734
	Netrako	268	600	1105	1712	378		4063
	Rajbari	284	424	499	443	218		1868
	Shariat	450	500	519	606	238		2313
	Sherpur	297	713	880	750	285		2925
	Tangail	1608	1702	1514	1300	931		7055
Khulna	Bagerha	932	751	487	389	417		2976
	Chuadan	176	328	504	528	317	267	2120
	Jessore	926	978	1065	995	279	506	4749
	Jhenaid	350	602	721	926	227	223	3049
	Khulna	592	500	476	466	1174	1167	4375
	Kushtia	1278	738	663	524	186	253	3642
	Magura	205	330	367	490	216		1608
	Meherpu		448	348	312	150		1258
	Narail	310	334	276	273	147		1340
	Satkhir	824	840	810	824	138	156	3592

Rajshahi	Bogra	1967	1376	1114	999	233	531	6220
	Dina	1050	1216	1273	1202	307	460	5508
	Gaiba	1109	1490	1154	873	487		5113
	Joypu	234	424	486	438	255		1837
	Kurig	504	986	967	847	606		3910
	Lalmo	272	482	596	711	353		2414
	Naoga	578	911	1336	1912		403	5140
	Nator	546	615	725	842		413	3141
	Nawab	755	533	484	352		520	2644
	Nilph	412	688	731	788	225	249	3093
	Pabna	1261	702	654	659	449	407	4132
	Panch	274	373	486	449	153		1735
	Rajshahi	551	802	880	813	717	1009	4772
	Rangp	1090	1400	1202	950	556	382	5580
	Siraj	2490	997	692	782	685		5646
	Thaku	288	376	622	783	214		2283
Sylhet	Habigan	400	636	909	1037	264	101	3347
	Maulvib	1109	792	607	324	146	153	3131
	Sunamga	292	629	1002	1455	290	91	3759
	Sylhet	1445	1193	729	433	213	568	4581
Bangladesh		48032	47496	47471	47101	32078	26726	248904

Appendix 4

2009 Master Sample

Sample PSU Count by Division, Zila and Urban/Rural Classification

Division	Zila	Rural				Urban		Total
		1	2	3	4	1	2	
Barisal	Barguna	3	4	5	6	2		20
	Barisal	15	11	9	8	4	5	52
	Bhola	5	6	9	11	4	1	36
	Jhaloka	5	5	3	1	1	1	16
	Patuakh	7	7	7	7	2	2	32
	Pirojpu	7	5	6	4	2	2	26
Chittagong	Bandarb	1			1	1		3
	Brahman	3	5	6	7	2	1	24
	Chandpu	4	7	6	3	4		24
	Chittagong	13	8	7	4	13	23	68

Dhaka	Dhaka	4	1	2		24	50	81
	Faridpu	3	3	4	3	2		15
	Gazipur	4	3	2	2	6	3	20
	Gopalga	3	2	2	2	1		10
	Jamalpu	2	5	6	6	3		22
	Kishorg	5	5	5	6	3		24
	Madarip	2	2	3	2	1		10
	Manikga	4	3	3	2	1		13
	Munshig	6	2	2		1		11
	Mymensi	5	9	12	11	5	1	43
	Narayan	8	1			7	5	21
	Narsing	6	4	2	2	3		17
	Netrako	1	3	5	8	2		19
	Rajbari	1	2	2	2	1		8
	Shariat	2	2	2	3	1		10
	Sherpur	1	3	4	4	1		13
	Tangail	8	8	7	5	5		33
Khulna	Bagerha	7	6	3	3	3		22
	Chuadan	1	3	3	4	2	2	15
	Jessore	7	8	8	7	2	3	35
	Jhenaid	3	4	6	7	1	2	23
	Khulna	4	4	4	3	9	10	34
	Kushtia	8	6	4	4	1	2	25
	Magura	2	2	2	4	1		11
	Meherpu		3	3	2	1		9
	Narail	3	2	2	2	1		10
	Satkhir	6	6	6	6	1	1	26
Rajshahi	Bogra	11						

References

1. Cochran, W.G. (1963) *Sampling Techniques*, New York: Wiley

2. Elbers, C., J. Lanjouw and P. Lanjouw. (2003) Micro-Level Estimation of Poverty and Inequality. *Econometrica*, 71(1), 355-364.

3. Kish, L. (1965) *Survey Sampling* , New York: Wiley

4. Kish, L. (1987) *Statistical Design for Research*, New York: Wiley

5. Kozak, M. (2004). Optimal stratification using random search method in agricultural surveys. *Statistics in Transition*, 6(5), 797-806.

6. Lohr, S.L. (2010) *Sampling: Design and Analysis*, Second edition, Boston: Brooks/Cole

7. Sethi, V. K. (1963) A note on optimum stratification of populations for estimating the population means. *The Australian Journal of Statistics*, **5**, 20-33.

8. United Nations Statistical Office (1950), *The Preparation of Sampling Survey Reports*, New York: U.N. Series C, No. 1

9. United Nations Secretariat (2005) *Household Sample Surveys in Developing and Transition Countries*, Publication number ST/ESA/STAT/SER.F/96, New York: U.N.

10. http://www.inside-r.org/packages/cran/stratification/docs/strata.LH accessed 11 June, 2013.