Column Description in Table A

Column 1 – Regions and provinces in the Philippines
Column 2 – Total household population based on Census 2000 counts
Column 3 – Number of households based on Census 2000 counts
Column 4 – Number of PSUs formed per region/province/city
Column 5 – Number of sample households allocated per region/province/city
Column 6 – Number of sample households allocated per region/province/city adjusted to cover for the non-response
Column 7 – Number of sample PSUs per region/province/city from which sample households will be drawn
Column 8 – Number of sample self-representing PSUs per region/province/city
Column 9 – Number of sample non-self-representing PSUs rounded off to the nearest multiple of four
Column 10 – Total of Columns 8 and 9

✓ *For additional information, please write, call or visit*

NATIONAL STATISTICS OFFICE
P.O. Box 779, Manila Philippines

Statistical Methodology Unit (SMU)
4/F Solicarel Bldg. I, R. Magsaysay Blvd.
Sta. Mesa, Manila
Tel. No.: 713-2789
E-mail: M.Lorenzo@mail.census.gov.ph

✓ *For a copy of The 2003 Master Sample Documentation, please contact*

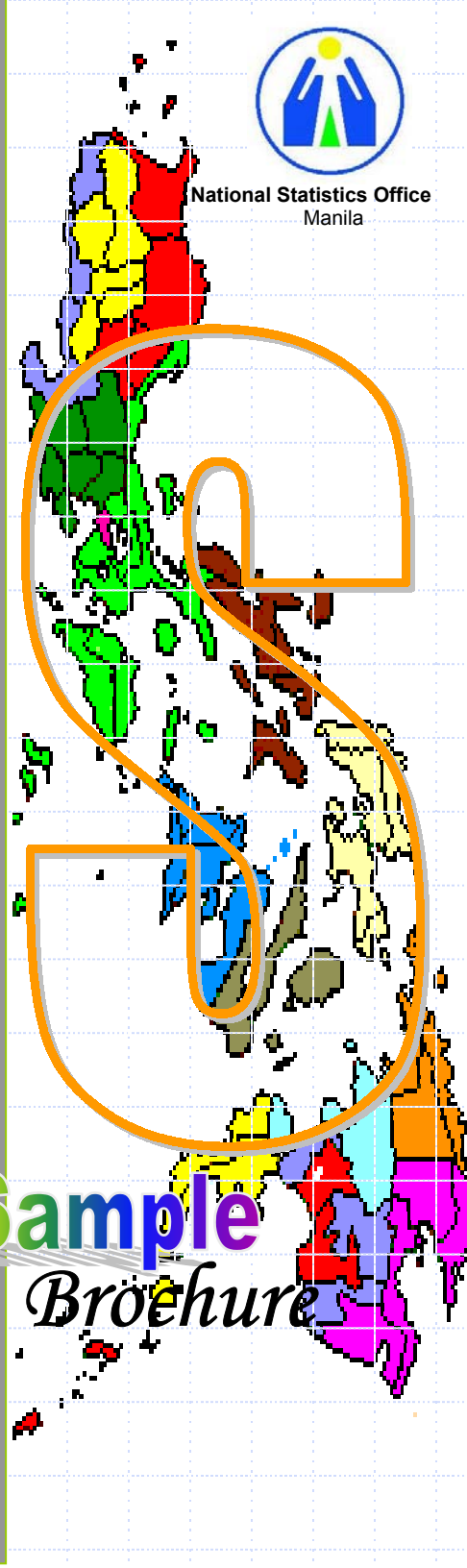
Databank and Information Services Division
2/F Soledad Bldg., R. Magsaysay Blvd.
Sta. Mesa, Manila
Tel. Nos.: 713-7081/715-6430
Fax: 714-1715

E-mail: info@mail.census.gov.ph
sales@mail.census.gov.ph

2003

Master Sample

Brochure



National Statistics Office
Manila

Beginning July 2003, the National Statistics Office (NSO) employs the 2003 master sample (MS) design in the conduct of its household surveys. The 2003 MS extensively employed the results of the 2000 Census of Population and Housing as well as results of past national surveys, such as the 2000 Family Income and Expenditure Survey (FIES), the 2001 Labor Force Survey (LFS), and the 1997 Family Planning Survey (FPS).

This note provides an overview and general description of the different aspects of the 2003 MS. More thorough discussions are given in the main technical documentation (The 2003 Master Sample Documentation).

A **master sample** is defined as a sample from which subsamples are drawn to serve the needs of several surveys. Master samples are usually employed for several surveys covering different themes that are integrated in terms of target population, sample design and field operations. The use of master samples promotes efficiency on the use of limited resources (e.g. single cost for the development of survey design and preparation of sampling frames). It also allows the linking of the different survey variables thereby creating a richer database for more meaningful and useful analyses. Usually, a master sample is an area sample of clusters of households referred to as Primary Sampling Units (PSUs).

With the availability of updated information for the general household population from the 2000 Census of Population and Housing, a redesign of the master sample was done.

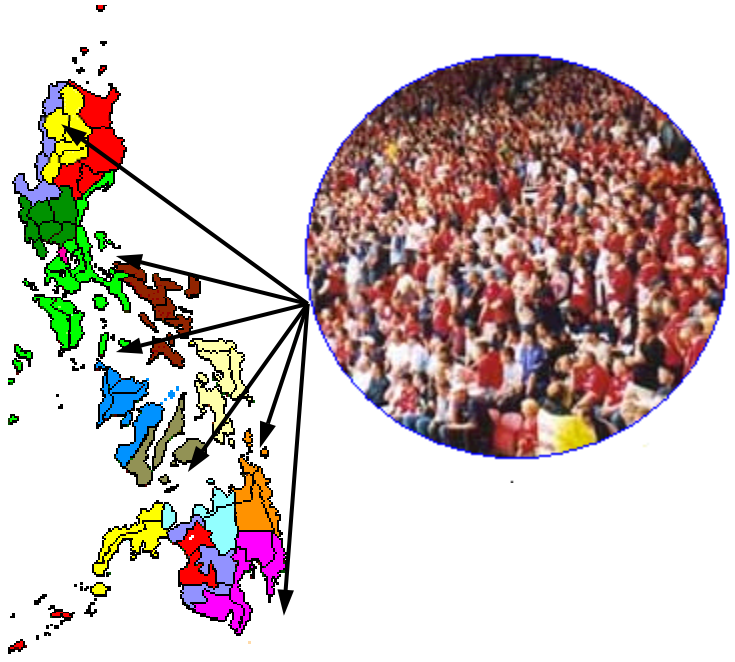
Region/ Province	Total Pop'n	No. Of Hhlds	No. Of PSU	Allocated Sample Size		No. Of Sample PSU	Final PSU Allocation		
				Original	Adj. For Non Response		SR PSU	NSR PSU	Total PSU
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Paranaque City	437,738	92,589	11	198	219	16	11	0	11
Pasay City	327,335	72,878	85	156	172	13	2	12	14
Pateros	57,109	12,098	9	26	29	2	1	4	5
Taguig	454,403	100,756	9	215	238	18	9	0	9
CAR	1,339,703	265,460	360	1,838	1935	115	7	108	115
Abra	209,791	41,054	58	284	299	18	0	16	16
Benguet	326,688	64,833	74	449	473	28	5	24	29
Ifugao	153,643	30,117	49	209	220	13	0	12	12
Kalinga	170,890	30,475	43	211	222	13	1	12	13
Mountain Province	132,795	26,703	40	185	195	12	0	12	12
Apayao	92,743	17,542	29	121	128	8	0	8	8
Baguio City	253,153	54,736	67	379	399	24	1	24	25
ARMM	3,073,420	496,256	674	2,013	2115	126	2	124	126
Basilan	210,504	40,461	53	164	172	10	0	8	8
Lanao del Sur	749,325	108,711	177	441	463	28	0	28	28
Maguindanao	947,918	163,297	196	663	696	41	1	40	41
Sulu	696,427	106,292	141	431	453	27	1	28	29
Tawi-Tawi	339,957	57,343	75	233	244	15	0	16	16
Marawi City	129,289	20,152	32	82	86	5	0	4	4
CARAGA	2,083,590	397,955	502	1,928	2053	120	8	112	120
Agusan del Norte	285,755	53,506	75	259	276	16	0	16	16
Agusan del Sur	551,212	103,621	131	502	535	31	3	28	31
Surigao del Norte	476,597	93,517	122	453	482	28	2	24	26
Surigao del Sur	502,700	95,706	118	464	494	29	2	28	30
Butuan City	267,326	51,605	56	250	266	16	1	16	17

Region/ Province	Total Pop'n	No. Of Hhlds	No. Of PSU	Allocated Sample Size		No. Of Sample PSU	Final PSU Allocation		
				Original	Adj. For Non Response		SR PSU	NSR PSU	Total PSU
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Iligan City	284,503	57,207	37	183	192	11	1	12	13
Cagayan de Oro City	482,310	98,131	32	314	330	20	10	8	18
REGION 11	3,666,787	754,218	657	2,300	2525	144	23	120	143
Davao del Norte	743,592	150,627	139	459	504	29	3	24	27
Davao del Sur	745,401	154,484	168	471	517	29	1	28	29
Davao Oriental	448,234	87,200	89	266	292	17	1	16	17
Compostella	579,609	120,805	129	368	404	23	3	20	23
Davao City	1,149,951	241,102	132	735	807	46	15	32	47
REGION 12	3,230,852	646,668	648	2,171	2386	136	17	120	137
Cotabato	965,698	190,005	229	638	701	40	3	36	39
South Cotabato	689,703	141,230	140	474	521	30	3	28	31
Sultan Kudarat	587,643	114,381	130	384	422	24	1	24	25
Sarangani	410,221	82,804	96	278	305	17	1	16	17
Cotabato City	166,477	31,724	28	106	117	7	0	8	8
General Santos City	411,110	86,524	25	290	319	18	9	8	17
NCR	9,570,589	2,066,392	982	4,413	4882	368	193	164	357
Manila City	1,546,711	326,869	408	698	772	58	4	56	60
Mandaluyong City	265,222	57,871	26	124	137	10	5	4	9
Marikina City	371,663	76,272	9	163	180	14	9	0	9
Pasig City	508,084	107,960	28	231	255	19	15	4	19
Quezon City	2,075,912	459,989	130	982	1,087	82	48	36	84
San Juan	115,124	23,422	18	50	55	4	1	4	5
Caloocan City	1,138,788	242,436	128	518	573	43	22	20	42
Malabon	328,774	72,607	21	155	172	13	8	4	12
Navotas	222,928	48,085	13	103	114	9	7	4	11
Valenzuela City	476,969	105,444	29	225	249	19	13	8	21
Las Pinas City	440,315	92,203	20	197	218	16	14	4	18
Makati City	453,881	100,678	31	215	238	18	17	4	21
Muntinlupa City	349,633	74,235	7	159	175	13	7	0	7

Target Population

The 2003 master sample design covers all households in the Philippines excluding institutional households as well as households in the Least Accessible Barangays (LABS).

For the 2003 MS, a barangay is classified as LAB if: (a) there is no regular means of transportation (frequency of transportation is less than three times a week); (b) the cost of a one-way fare is more than 500 pesos; or, (c) it takes more than 8 hours of walking to reach the barangay. The LABS were identified by the NSO field offices. The final list was determined after further consultation by the NSO Central Office MS project team with the NSO field offices. A total of 350 barangays were classified as LABs and were excluded in the MS frame.



Primary Sampling Units (PSU)

A master sample is a sample of PSUs. A PSU, on the other hand, is a cluster of households with clear and stable boundaries, that is, the boundaries do not change rapidly over time. A PSU should also contain sufficient number of households to support all the household surveys for which it will be used as sample. The 2003 MS for instance, needs PSUs with at least 500 households.

The barangays were found to be the most suitable administrative unit (in terms of number) to form the PSUs for the 2003 MS. However, more than half of the barangays do not satisfy the minimum size requirement (number of household) of an ideal PSU, thus, "small" barangays were grouped with contiguous barangays within the municipality to form the desired PSUs.

Do you know that…

- There are 41,942 barangays in the country, 350 of which were considered least accessible barangays (LABs) and were excluded from the frame
- The total number of PSUs formed from 41,592 barangays is 16,586
- The average number of households in a 2003 MS PSU (or PSU size) is 923

A list of all the PSUs formed and their characteristics in terms of the stratification variables used is contained in the Master Sample Frame (MSF).

Region/ Province	Total Pop'n	No. Of Hhlds	No. Of PSU	Allocated Sample Size		No. Of Sample PSU	Final PSU Allocation		
				Original	Adj. For Non Response		SR PSU	NSR PSU	Total PSU
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
REGION 7	5,694,537	1,142,038	1,247	2,848	3046	178	9	168	177
Bohol	1,129,095	213,879	320	533	570	33	0	32	32
Cebu	2,378,932	478,026	553	1,192	1,275	74	2	72	74
Negros Oriental	1,127,621	228,272	251	569	609	36	0	36	36
Siquijor	81,149	17,324	28	43	46	3	0	4	4
Cebu City	715,424	148,915	69	371	397	23	7	16	23
Mandaue City	262,316	55,622	26	139	148	9	0	8	8
REGION 8	3,577,761	712,715	1,048	2,249	2524	141	0	140	140
Eastern Samar	375,078	73,646	117	232	261	15	0	16	16
Leyte	1,430,081	291,039	421	918	1,031	57	0	56	56
Northern Samar	485,265	91,874	129	290	325	18	0	16	16
Samar	626,633	121,965	182	385	432	24	0	24	24
Southern Leyte	358,702	72,930	115	230	258	14	0	12	12
Biliran	139,379	27,974	42	88	99	6	0	8	8
Ormoc City	162,623	33,287	42	105	118	7	0	8	8
REGION 9	2,807,013	547,407	675	2,064	2152	129	11	120	131
Zamboanga del Norte	809,672	159,463	222	601	627	38	0	40	40
Zamboanga del Sur	831,504	161,751	222	610	636	38	0	40	40
Zamboanga Sibugay	495,539	93,910	139	354	369	22	0	20	20
Isabela City	72,319	13,959	18	53	55	3	0	4	4
Zamboanga City	597,979	118,324	74	446	465	28	11	16	27
REGION 10	3,546,819	698,505	785	2,232	2349	139	12	128	140
Bukidnon	1,058,283	201,272	228	643	677	40	1	40	41
Camiguin	74,605	15,052	20	48	51	3	0	4	4
Lanao del Norte	476,106	90,675	136	290	305	18	0	16	16
Misamis Occidental	497,004	101,971	154	326	343	20	0	20	20
Misamis Oriental	674,008	134,197	178	429	451	27	0	28	28

Region/ Province	Total Pop'n	No. Of Hhlds	No. Of PSU	Allocated Sample Size		No. Of Sample PSU	Final PSU Allocation		
				Original	Adj. For Non Response		SR PSU	NSR PSU	Total PSU
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cavite	2,090,786	436,356	459	942	979	59	5	56	61
Laguna	1,972,247	419,163	350	905	941	57	6	52	58
Quezon	1,473,460	298,778	399	645	671	40	0	40	40
Rizal	1,746,603	364,886	126	788	819	49	20	28	48
Lucena City	191,504	38,958	23	84	87	5	1	4	5
MIMAROPA	2,253,006	452,790	596	1,974	2052	123	2	124	126
Marinduque	216,887	43,892	67	191	199	12	0	12	12
Occidental Mindoro	368,210	74,420	84	324	337	20	1	20	21
Oriental Mindoro	676,651	133,971	191	584	607	36	0	36	36
Palawan	728,723	147,069	179	641	666	40	1	40	41
Romblon	262,535	53,438	75	233	242	15	0	16	16
REGION 5	4,659,730	892,720	1,242	2,483	2667	155	0	156	156
Albay	1,083,327	208,039	300	579	621	36	0	36	36
Camarines Norte	465,098	90,982	126	253	272	16	0	16	16
Camarines Sur	1,408,937	261,686	383	728	782	45	0	44	44
Catanduanes	215,616	41,109	61	114	123	7	0	8	8
Masbate	709,737	140,458	199	391	420	24	0	24	24
Sorsogon	644,364	124,944	152	347	373	22	0	24	24
Naga City	132,651	25,502	21	71	76	4	0	4	4
REGION 6	6,227,183	1,220,660	1,503	2,970	3282	186	7	176	183
Aklan	456,822	89,375	125	217	240	14	0	12	12
Antique	454,149	90,186	135	219	242	14	0	12	12
Capiz	652,955	128,479	183	313	345	20	0	20	20
Iloilo	1,554,030	298,585	468	726	803	45	0	44	44
Negros Occidental	2,163,915	423,839	420	1,031	1,140	64	0	64	64
Guimaras	140,741	27,496	39	67	74	4	0	4	4
Iloilo City	363,706	72,459	91	176	195	11	0	12	12
Bacolod City	440,865	90,241	42	220	243	14	7	8	15

Domains

Survey estimates are generally needed for the nation as a whole as well as for various subgroups. These subgroups may refer to socio-demographic subdivisions that are usually spread throughout the population such as female-headed households by age of head or educational levels by age and sex, or geographic subdivision such as regions or provinces. Thus, the survey may be designed taking into consideration the provision of estimates with adequate level of precision for such subdivisions. At the design stage, geographic subdivisions are usually treated as domains. A domain refers to such subdivisions in which estimates of adequate precision are desired.

Based on past surveys and other available resources, most national surveys are able to produce estimates of adequate precision at the regional level only. The precision of estimates may be measured in several ways. One way is to construct a 95% confidence interval estimate (note that a wider confidence interval estimate is deemed imprecise and less useful).

Example: The estimated proportion of poor families for a given domain is 30%

Coefficient of Variation (CV)	Standard Error (SE)	95% Confidence Interval Estimate
10%	3%	30% ± (2*3%) ⇒ 24% to 36%
20%	6%	30% ± (2*6%) ⇒ 18% to 42%

The example above means that with a CV of 10%, the true proportion of poor families lies between 24% to 36% ninety-five percent of the time. A CV of 20%, on the other hand, assures that the true proportion of poor families lies between 18% to 42% ninety-five percent of the time. Notice that the width of the interval widens as the CV or SE values increases. A summary of the provincial and regional level CV values of the estimated proportion of poor families is shown in Table 1.

Table 1. Distribution of Regional and Provincial Estimates of the Proportion of Poor Families Based on the Results of the 2000 Family Income and Expenditures Survey (FIES)

Range of CV Values	Number of Regional Estimates	%
< 5%	6	35.3
5 % - 10%	11	64.7
Total	17	100.0

Range of CV Values	Number of Provincial Estimates	%
< 5%	3	3.7
5 % - 10%	36	43.9
10% - 15%	33	40.2
15% - 20%	8	9.8
20% - 25%	1	1.2
> 25%	1	1.2
Total	82	100.0

Source of Primary Data: NSO, 2000 FIES

For domain specification, an estimate is considered precise if the CV value of the estimated proportion of poor families does not exceed 10%. This criterion was used in specifying regions as domains of the MS. Note that in Table 1, only 39 out of 82 provincial estimates of the proportion of poor households yielded CV values less than 10%.

The importance of generating provincial level estimates was seriously considered in defining major sampling domains for the MS. However, generating provincial level estimates with adequate precision requires larger sample size that is usually not feasible and sustainable given the resources available for the survey.

Table A

Sample size distribution by region and province. 2003 NSO Master Sample.

Region/ Province	Total Pop'n	No. Of Hhlds	No. Of PSU	Allocated Sample Size		No. Of Sample PSU	Final PSU Allocation		
				Original	Adj. For Non Response		SR PSU	NSR PSU	Total PSU
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PHILIPPINES	76,311,169	15,312,424	16,579	43,882	46,976	2,835	330	2,496	2,826
REGION 1	4,192,048	837,348	1,199	2,408	2543	150	0	148	148
Ilocos Norte	513,850	108,477	164	312	329	19	0	20	20
Ilocos Sur	589,797	119,270	197	343	362	21	0	20	20
La Union	655,561	131,140	189	377	398	24	0	24	24
Pangasinan	2,432,840	478,461	649	1,376	1,453	86	0	84	84
REGION 2	2,776,100	568,347	839	2,085	2240	130	0	132	132
Batanes	16,548	3,489	5	13	14	1	0	48	48
Cagayan	969,824	196,046	292	719	773	45			
Isabela	1,172,502	239,624	362	879	944	55	0	56	56
Nueva Vizcaya	362,603	75,920	109	279	299	17	0	16	16
Quirino	144,203	29,904	46	110	118	7	0	8	8
Santiago City	110,420	23,364	25	86	92	5	0	4	4
REGION 3	8,228,567	1,676,713	1,780	3,726	3882	233	7	224	231
Bataan	556,930	113,596	135	252	263	16	0	16	16
Bulacan	2,235,626	465,743	420	1,035	1,078	65	1	64	65
Nueva Ecija	1,659,257	342,216	447	761	792	48	0	48	48
Pampanga	1,629,273	310,483	309	690	719	43	3	40	43
Tarlac	1,077,289	217,940	259	484	505	30	1	28	29
Zambales	431,625	91,269	115	203	211	13	0	12	12
Aurora	172,963	34,896	50	78	81	5	0	4	4
Angeles City	271,383	57,367	30	127	133	8	1	8	9
Olongapo City	194,221	43,203	15	96	100	6	1	4	5
CALABARZON	9,383,464	1,936,232	1,842	4,181	4346	261	32	232	264
Batangas	1,908,864	378,091	485	816	849	51	0	52	52

Table 2. Sample rotation design from 2004 to 2008

Year	Quarter	Sample/Rotation Cluster*	
2004	January	A1	B1
	April	A2	B2
	July	A3	B3
	October	A4	B4
2005	January	A1	B5
	April	A2	B6
	July	A3	B7
	October	A4	B8
2006	January	A5	B5
	April	A6	B6
	July	A7	B7
	October	A8	B8
2007	January	A7	B7
	April	A6	B9
	July	A5	B10
	October	A8	B11
2008	January	A9	B12
	April	A10	B9
	July	A11	B10
	October	A12	B11

* Numbers represent rotation groups formed for the housing units withn the sampled EAs and letters represent rotation clusters. Rotation cluster A includes replicates one and two while rotation cluster B includes replicates 3 and 4.

With regions as domains, the computed total sample size that would give the desired reliability in the estimates for each domain is manageable. In particular, the required sample size per region was computed so that the expected CV of the estimated proportion of poor households would not exceed 5% except in the NCR where the CV value was set to 10%. The exception was made through the observation that the estimated proportion of poor households in NCR is small (around 8%). The total sample size computed that satisfies this reliability condition is about 43,000 households. If provinces were to be specified as domains, the total sample size requirement would be much larger than this.

Sample Allocation

The procedure in allocating the total sample size in each domain directly affects the precision of the estimates based on two important purposes. These are:

- The need to generate precise estimates at the national level or subclasses of the population that cuts across domains. Examples of subclass estimates are the proportion of poor households among female-headed households or the employment rate by major industry classification (e.g. agriculture, manufacturing, etc.). For this purpose, allocating the sample proportional to the total number of households in the domain is considered the best solution.

- The need to generate precise estimates at the domain level for purposes of comparison. In this case, allocating the total sample size equally across domains is the best solution.

Clearly, the best solutions for each of the two concerns are not consistent with one another. Because of this, a compromise allocation scheme was used. In particular, the Kish Allocation Scheme was used to allocate the total sample size to each domain.

The final sample size per region was further adjusted (increased) to consider projected non-response and population growth. These adjustments resulted to a total sample size of about 47,000 households.

Under the Kish Allocation Scheme, the sample size in each domain, denoted by n_d, is determined by

$$n_d = n\left[\sqrt{H^{-2}+W_d^2}\Big/\sum_{d=1}^{H}\sqrt{H^{-2}+W_d^2}\right] \qquad \text{Equation 1}$$

where:

n - total sample size (about 43,000);
H - number of specified domains/regions (=17); and
$W_d = N_d / N$ - proportion of the total household population (N) found in region d.

Note that Equation 1 gives equal importance to the two allocation concerns mentioned.

Number of PSUs per Domain/Region

The number of sampled PSUs per domain was computed by simply dividing the total sample size by the desired sample size per PSU. The desired sample size per PSU was determined using: (1) the information on the cost of data collection efforts in the region; and, (2) the indication of similarity or homogeneity of the households within the PSU. The basic idea is to take smaller samples with PSUs consisting of homogeneous households and if the cost of data collection is more expensive. With these information gathered from past survey results, the number of sample households from each PSU was set at 16 for areas outside the National Capital Region (NCR) and 12 for the NCR. This means that for NCR, the total number of PSUs is equal to the allocated sample size divided by 12. For the other regions, it is equal to the allocated sample size divided by 16.

Selection of Subsamples

Several options are available in the selection of subsamples from the new master sample. These options depend on whether the survey is done together with the regular Labor Force Survey (LFS) or as a stand-alone survey.

- If a survey that requires only a subsample is conducted together with the LFS, then it is more efficient to select a subsample of housing units within a PSU. For instance, suppose the total number of sampled housing units within a PSU is 16, a quarter sample is drawn by selecting 4 housing units from among the 16 with equal probability.

- If the survey is to be conducted independently of the LFS, then it is more efficient to select a subsample of PSUs rather than a subsample of housing units in all PSUs. The subsampling of PSUs can be done by selecting one or more replicates. For instance, if a 50% sample is desired, then this can be achieved by selecting two replicates. This applies on both SR and NSR PSUs.

Estimation Procedure

The generation of the survey weights for each responding element is one of the key activities in generating estimates using the MS. The weight may be interpreted as the relative importance given to the responding unit in the generation of estimates. This can also be interpreted as the number of non-sampled units that each responding unit represents in the sample. Basically, the **final survey weight** is defined as the product of: (1) Base weights; (2) The nonresponse adjustment weight; and, (3) Weight adjustment based on known population totals or simply post-stratification weight. The base weight is determined by taking the inverse of the selection probabilities of each unit of analysis. The nonresponse adjustment weight is determined by taking the inverse.

Rotation of Samples

The MS will be used for a period of 10 years. As such, sample elements need to be replaced by a new set at certain points in time. Retaining the original sample elements would create problems such as response burden that would eventually affect the overall quality of the survey results. In addition, units repeatedly interviewed increase the likelihood of non-response. A solution to this problem is to devise a sample rotation plan so that a unit may stay in the sample for some period and then replaced permanently by a new set of sample. To facilitate a sample replacement scheme, each replicate will form a panel. In each PSU, all units were divided into rotation groups of equal size. The sample replacement scheme is such that every quarter of the year, a new rotation group in each panel will be selected. However to maximize the effect of the correlation of the estimates between years, 50% of the panels will have common samples for a quarter in consecutive years. For illustration, refer to the proposed sample rotation design in Table 2.

Future Direction

The completion of the research for 2003 master sample design directed the NSO, through the Statistical Methodology Unit (SMU), to conduct other related research studies. For 2004, the research study line up is as follows:

- Validation of Raking Procedure used for LFS Estimates;
- Provincial Estimation of Unemployment Using Aggregated Four Quarter Samples;
- Comparison of Estimates (levels/rates and precision) Using Old and New Nonresponse Adjustment Procedure; and
- Comparison of the number of households obtained in C2K and CA/CF listing by EA.

The final number of sample PSUs for each domain was determined by first classifying PSUs as either self-representing (SR) or non-self-representing (NSR). In addition, to facilitate the selection of subsamples, the total number of NSR PSUs in each region was adjusted to make it a multiple of 4.

The 2003 MS consist of a sample of 2,826 PSUs. The sample size distribution across regions and provinces are shown in the attached Table A.

Definition

SR PSU or Self-Representing Primary Sampling Unit – a very large PSU in the region/domain with a selection probability of approximately 1 or higher and is outright included in the MS; it is properly treated as a stratum; also known as certainty PSU

NSR PSU or Non-Self-Representing Primary Sampling Unit – a regular to small sized PSU in a region/domain; also known as non-certainty PSU

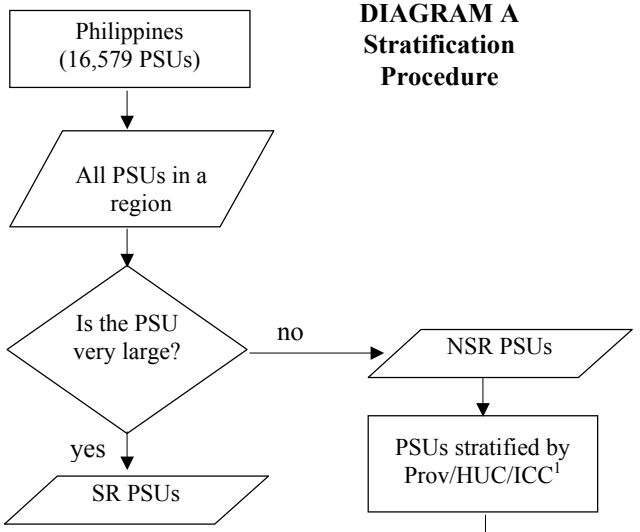
Stratification of PSUs

Stratification involves the division of the entire population into non-overlapping subgroups called strata, from which samples are being selected independently. This procedure is done to:

- Improve the efficiency of the estimates as a result of combining units that are similar in characteristic. This means improving on the precision of the estimates for a given sample size.
- Provide samples for specific subgroups of the population in which separate estimates are desired.

The stratification procedure used in the 2003 MS is described in Diagram A.

DIAGRAM A
Stratification Procedure



SR PSUs are treated as separate strata

❖ A total of 955 explicit strata were formed, 330 of which were the SR PSUs.

[1] This allows the generation of either direct or indirect sub-regional estimates

[2] Proportion of strongly built houses

[3] An indication of the proportion of households engaged in agriculture

[4] Per capita municipal income

Sample Selection

In each explicit stratum, a sample of PSUs, and then sample EAs within PSUs, was selected with probability proportional to size (PPS) where size is the number of households enumerated in the 2000 Census of Population and Housing (CPH). Within each sampled EA, a sample of housing units was selected with equal probability. All households in the housing units sampled are completely enumerated, except for few cases when the housing units have more than three households. For operational considerations, the maximum number of household that could be enumerated in each sampled housing units is three. In the case of SR PSUs, the EAs were the PSUs and a minimum of two EAs were selected with PPS to ensure valid estimation of the variances.

Formation of Replicates

Another important feature of the 2003 MS design is its flexibility to meet the needs of different surveys. Some surveys require only a fewer set of sample and thus the need to sub-sample from the master sample. To facilitate the selection of sub-samples, the MS was divided into four replicates. A **replicate** is defined as a subsample that possesses the properties of the full master sample such that each replicate is able to generate national level estimates of adequate precision.

For the NSR PSUs, each of the four PSUs in every stratum is assigned to one replicate. In the case of SR PSUs, on the other hand, the EAs were distributed to the replicates in such a way that a balance between two half samples (each of two replicates) can be achieved. A balanced distribution of EAs of the SR PSUs to the four replicates can not be achieved because most of the SR PSUs have only two EAs.