SURVEY DESIGN

A. Population Coverage

The 2006 FIES has as its target population, all households and members of households nationwide. A household is defined as an aggregate of persons, generally but not necessarily bound by ties of kinship, who live together under the same roof and eat together or share in common the household food. Household membership comprises the head of the household, relatives living with him such as his or her spouse, children, parent, brother or sister, son-in-law or daughter-in-law, grandson or granddaughter and other relatives. Household membership likewise includes boarders, domestic helpers and non-relatives. A person who lives alone is considered a separate household.

Excluded in the target population are households in the least accessible barangays (LABs). A barangay is classified as LAB if: (a) it requires more than eight hours walk from the last vehicle station; and or, (b) the frequency of transportation is less than three times a week and the cost of a one-way trip is more than five hundred pesos. A total of 350 barangays were classified as LABs. This number accounts for only 0.83 percent of the total number of barangays in the country. The total number of households in these areas account for only 0.38 percent of the total number of households.

B. Sampling Design

The 2006 FIES used the sampling design of the 2003 Master Sample (MS) for household surveys starting in July 2003.

1. Domain

The 2003 MS considers the country's 17 administrative regions as defined in Executive Orders (EO) 36 and 131 as the sampling domains. A domain is referred to as a subdivision of the country for which estimates with adequate level of precision are generated. It must be noted that while there is a demand for data at the provincial level (and to some extent municipal and barangay levels), the provinces were not treated as sampling domains because there are more than 80 provinces which would entail a large resource requirement. Below are the 17 administrative regions of the country:

National Capital Region
Cordillera Administrative Region
Region I - Ilocos
Region II – Cagayan Valley
Region III – Central Luzon
Region IVA – CALABARZON
Region IVB – MIMAROPA
Region V – Bicol
Region VI – Western Visayas
Region VII – Central Visayas
Region VIII – Eastern Visayas
Region IX – Zamboanga Peninsula
Region X – Northern Mindanao
Region XI – Davao
Region XII – SOCCSKSARGEN
Region XIII – Caraga
Autonomous Region in Muslim Mindanao

2. Sampling Frame

As in most household surveys, the 2003 MS made use of an area sample design. For this purpose, the Enumeration Area Reference File (EARF) of the 2000 Census of Population and Housing (CPH) was utilized as sampling frame. The EARF contains the number of households by enumeration area (EA) in each barangay.

This frame was used to form the primary sampling units (PSUs). With consideration of the period for which the 2003 MS will be in use, the PSUs were formed or defined as a barangay or a combination of barangays with at least 500 households.

3. Stratification

The 2003 MS considers the 17 regions of the country as the primary strata. Within each region, further stratification was performed using geographic groupings such as provinces, highly urbanized cities (HUCs), and independent component cities (ICCs). Within each of these substrata formed within regions, the PSUs were further stratified, to the extent possible, using the proportion of strong houses (PSTRONG), indicator of engagement in agriculture of the area (AGRI), and a measure of per capita income (PERCAPITA) as stratification factors.

of either galvanized iron, aluminum, concrete or clay tile, half galvanized-half concrete, or asbestos. The outer wall is considered made of strong materials if it is made of concrete, brick, stone, wood, half concrete-half wood, galvanized iron, asbestos or glass.

AGRI was determined in the following way: initially, an indicator variable was computed at the barangay level. That variable has a value of 1 if more than 50 percent of the households in the barangay were engaged in agriculture or fisheries and 0 otherwise, based on the 2000 CPH Barangay Schedule. To obtain a measure at the PSU level, a weighted average of the barangay indicator variable was computed for all the barangays within the PSU, weighted by the total number of households in the barangay. Thus, the value of AGRI at the PSU level lies between 0 and 1.

PERCAPITA is defined as the total income of the municipality divided by the total population in that municipality. Note that the PERCAPITA value of the PSUs is the same if the PSUs belong to the same municipality. The data on municipal income refer to year 2000 and were taken from the Department of Finance. However, if the 2000 municipal income was not reported to the Bureau of Local Government Finance (BLGF), 2001 income was used. If no 2000 or 2001 municipal income was reported, the income classification from the BLGF for this municipality was obtained. Using the data on municipal income, which are presented in income intervals, the average of the lower and the upper values of the income interval for the municipal class to which this municipality belongs were determined.

4. Sample Selection

The 2003 MS consists of 2,835 PSUs. The entire MS was divided into four sub-samples or independent replicates, such as a quarter sample contains one-fourth of the total PSUs; a half sample contains one-half of the four sub-samples or equivalent to all PSUs in two replicates.

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 sub-samples, the total number of NSR PSUs in each region was adjusted to make it a multiple of 4.

SR PSUs refers to 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 PSUs. NSR PSUs refers to a regular too small sized PSU in a region/domain; also known as non certainty PSUs. The 2003 MS consists of 330 certainty PSUs and 2,505 non-certainty PSUs.

To have some control over the sub-sample size, the PSUs were selected with probability proportional to some estimated measure of size. The size measure refers to the total number of households from the 2000 CPH. Because of the wide variation in PSU sizes, PSUs with selection probabilities greater than 1 were identified and were included in the sample as certainty selections.

At the second stage, enumeration areas (EAs) were selected within sampled PSUs, and at the third stage, housing units were selected within sampled EAs. Generally, all households in sampled housing units were enumerated, except for few cases when the number of households in a housing unit exceeds three. In which case, a sample of three households in a sampled housing unit was selected at random with equal probability.

An EA is defined as an area with discernable boundaries within barangays consisting of about 150 contiguous households. These EAs were identified during the 2000 CPH. A housing unit, on the other hand, is a structurally separate and independent place of abode which, by the way it has been constructed, converted, or arranged, is intended for habitation by a household.

5. Sample Size

The 2006 FIES involved the interview of a national sample of about 51,000 sample households deemed sufficient to gather data on family income and family expenditure and related information affecting income and expenditure levels and patterns in the Philippines at the national and regional level. The sample households covered in the survey were the same households interviewed in the July 2006 and January 2007 round of the LFS.

C. Estimation Procedure

In the 2003 Master Sample Design, the probability that a household is included in the sample varies across domains/regions. However, the sampling design is *epsem* within domain (i.e. equal selection probabilities within region). The initial step in the construction of weights is to determine the unit's base weight. This is defined as the inverse of its selection probabilities. The base weight is further adjusted to take into account possible non-response and possibly to make the estimates conform to some known population totals.

1. Base Weights

In general, the base weight assigned to a sampled unit is the inverse of its selection probability. In particular, the base weight is computed as the inverse of equations 1 Non Self Representing (NSR) and 2 Self Representing (SR) below:

$$P(h\alpha\beta\gamma) = \frac{a_{h\alpha} M_{h\alpha}}{\Sigma_{h\alpha} M_{h\alpha}} \cdot \frac{M_{h\alpha\beta}}{M_{h\alpha}} \cdot \frac{C_{h\alpha}}{M_{h\alpha\beta}} \cdot \frac{k_{h\alpha\beta\gamma}}{K_{h\alpha\beta\gamma}} = f_d = \frac{n_d}{N_d} \qquad (1)$$

$$P(h\alpha\beta\gamma) = \frac{b_{h\alpha} M_{h\alpha\beta}}{M_{h\alpha}} \cdot \frac{C_{h\alpha}}{M_{h\alpha\beta}} \cdot \frac{k_{h\alpha\beta\gamma}}{K_{h\alpha\beta\gamma}} = f_d = \frac{n_d}{N_d} \qquad (2)$$

where:

h	stratum index
α	index denoting the PSU
β	index denoting the EA
γ	index denoting the household (HH)
d	index denoting the domain/region
n_d	total sample size allocated to region d
N_d	total number of households in region d
$f_d = n_d / N_d$	overall sampling fraction for region d
$M_{h\alpha}$	total number of HHs for the αth PSU in stratum h
$M_{h\alpha\beta}$	total number of HHs in the βth EA from the αth PSU in stratum h
$a_{h\alpha}$	total number of sample PSUs from stratum h, a = 1 for NSR PSU
$C_{h\alpha}$	total number of sample housing units for each sampled EA
$k_{h\alpha\beta\gamma}$	number of sampled households per housing unit with three as the maximum
$K_{h\alpha\beta\gamma}$	total number of households residing in a housing unit
b_{ha}	total number of EAs selected from αth PSU, b =1 for SR PSUs

That is, the base weight for NSR and SR samples are equal to equations 3 and 4, respectively:

$$w_1 = \frac{\sum_{h\alpha} M_{h\alpha}}{a_{h\alpha} M_{h\alpha}} \cdot \frac{M_{h\alpha}}{M_{h\alpha\beta}} \cdot \frac{M_{h\alpha\beta}}{C_{h\alpha}} \cdot \frac{K_{h\alpha\beta\gamma}}{k_{h\alpha\beta\gamma}} = \frac{N_d}{n_d} \qquad (3)$$

$$w_1 = \frac{M_{h\alpha}}{b_{h\alpha} M_{h\alpha\beta}} \cdot \frac{M_{h\alpha\beta}}{C_{h\alpha}} \cdot \frac{K_{h\alpha\beta\gamma}}{k_{h\alpha\beta\gamma}} = \frac{N_d}{n_d} \qquad (4)$$

Note that the last term will equal to 1.0 in cases when all households in the sampled housing unit are enumerated. That is, when households per housing unit do not exceed three.

2. Non-response Adjustments

All surveys experienced some degree of unit or total non-response in which a sampled and eligible unit fails to participate in the survey (for example, the unit may refuse to participate, or may never be at home at times the interviewer calls). Adjustments are made to the base weights to compensate for non-response by sampled units eligible for the survey. In essence, the adjustment inflates the base weights of "similar" responding units to compensate for each non-responding unit.

The most common form of non-response weighting adjustment is a weighting class adjustment and that is the type of adjustment being used for surveys based on the 2003 MS. The full sample of responding households and non-responding households is divided into a number of weighting classes or cells and non-response adjustment factors are computed for each cell c as

$$w_c^{'} = \frac{\sum_{i \in rc} w_{di} + \sum_{j \in mc} w_{dj}}{\sum_{i \in rc} w_{di}} = \frac{\sum_{i \in sc} w_{di}}{\sum_{i \in rc} w_{di}} \qquad (5)$$

The denominator of $w_c^{'}$ is the sum of the weights of responding households (indexed r) in cell c. The numerator adds together the sum of the weights for responding households and the sum of the weights for eligible non-responding households (indexed m for missing) in cell c. Together these two sums in the numerator give the sum of the weights for the total eligible sample (indexed s) in cell c. Thus, the non-response weight adjustment $w_c^{'}$ 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 whether are out of scope for a given survey) are excluded.

3. Population Weighting Adjustments

The basic weight for the sample households for the 2006 FIES can be expressed as an inverse of the selection probability described in equations 1 and 2. Hence, the base weight for a sampled household in domain *d* in a housing unit in which all households are included is

$$w_d = \frac{1}{f_d^{'}} \qquad (6)$$

The basic weight was adjusted to take into account household non-interviews, followed by an adjustment based on the household projections for the domain.

Generally, weighted sample distributions do not conform to known population distributions (e.g. projected population (person) counts or projected household counts). In particular, sample estimates of population (person) counts or household counts generally fall short of true population (person) or household counts because of non-coverage resulting from omission of units and from non-responses such as refusal, non-reachable areas (critical areas) and others.

Hence, further weighting adjustments, termed as population weighting adjustment, may be made to make the survey estimates based on the adjusted weight estimates consistent with known population distributions.

For adjusting household level estimates, the reference count of households is obtained by dividing the total projected population by the average household size. This is resorted to in the absence of projected number of households.

4. Final Survey Weight

The final survey weight assigned to each responding unit is computed as the product of the base weight, the non-response adjustment and the population weighting adjustment (in the case of households as responding unit – the household population weighting adjustment).

For FIES, a ratio of 1:1 is observed or one family to one household. Consider first the estimation of total households. Let y_i and w_i denote the value of variable y and the final weight for household i. The notation can be applied to persons or households or any other unit of analysis. The final weight w_i can be viewed as the number of households that household i represents; thus Σw_i estimates the total number of households in the country *N*. The survey estimate of the households total for variable y, denoted by Y, is then simply

$$\hat{Y} = \Sigma w_i y_i$$