SAMPLE DESIGN

Appendix C

Chapter 1, Section 1.7, in Volume I of this report contains basic information about the sample design and sample implementation for the 2005-06 National Family Health Survey (NFHS-3). This appendix contains additional details about the sampling, including the methodology for determining the sample size, the sampling frame, sample selection in rural and urban areas, stratification of the sample, the selection of households, and sample weights.

Sample Size

The determination of the overall sample size for NFHS-3 was governed by several considerations, including the magnitude of the key indicators, the subgroups for which the indicators are required, the desired level of precision of the estimates, the availability of resources, and logistical considerations. Besides the requirement of producing population and health indicators at both the national and state levels, two additional considerations were critical for determining the size of the sample for NFHS-3. The survey was tasked with producing a) HIV prevalence estimates at the national level and at the state level for each of the six states identified by the National AIDS Control Organization (NACO) as having high HIV prevalence (Andhra Pradesh, Karnataka, Maharashtra, Manipur, Nagaland, and Tamil Nadu), and Uttar Pradesh, and b) robust estimates of population and health indicators for slum areas and for non-slum areas in eight cities, namely Delhi, Chennai in Tamil Nadu, Hyderabad in Andhra Pradesh, Indore in Madhya Pradesh, Kolkata in West Bengal, Meerut in Uttar Pradesh, and Mumbai and Nagpur in Maharashtra. Both of these requirements were crucial for determining the target number of individual interviews. In light of these requirements, the sample size was estimated separately for the number of individual interviews with women and with men and the number of women and number of men to be tested for HIV.

Sample size for HIV tests

The expected level of HIV prevalence, the acceptable level of standard errors for HIV prevalence, and the design effect of the sample design were the three prime determinants of sample size for HIV tests at the national level and at the state level for each of the seven states for which separate HIV estimates were to be produced. To estimate the required sample size for HIV testing, it was necessary to make a reasonable assumption about the design effect of HIV prevalence for the NFHS-3 sample design. The design effect is defined as the ratio of the standard error of the HIV estimate for the NFHS-3 sample design to the standard error that would result if a simple random sample had been used. Based on the experience of testing HIV in national household surveys in other countries, a design effect of 1.25 was assumed. The actual design effect of HIV prevalence in NFHS-3 (1.19 for women and 1.33 for men) was very close to the assumed design effect of 1.25.

In the absence of any previous community-based estimate of HIV prevalence for India, the HIV prevalence rates as reported by NACO of 0.92 percent at the national level, 0.75 percent

for Tamil Nadu, and 1.25 percent for each of the remaining five high HIV prevalence states were used for estimation of the sample size for the number of HIV tests. For state-level estimates of HIV prevalence, the maximum permissible relative standard error for women and men together was set at 15 percent. However, for the national estimate, the maximum permissible relative standard error was set at a low level of 5 percent. The table below gives the assumptions at the national and state levels, the proposed sample sizes of women and men to be tested compatible with the assumptions, and the expected precision of the HIV prevalence estimate associated with the proposed sample sizes for India as a whole and for each of the six high HIV prevalence states.

Region	Assumed HIV prevalence rate	Assumed design effect	Proposed sample size	Estimated relative precision of resulting HIV prevalence estimate	Confidence interval	
					Lower bound	Upper bound
Tamil Nadu						
Women and men separately	0.75	1.25	6,400	18.0%	0.5%	1.0%
Women and men together	0.75	1.25	12,800	12.7%	0.6%	0.9%
Each of the remaining five high HIV prevalence states						
Women and men separately	1.25%	1.25	6,400	13.9%	0.9%	1.6%
Women and men together	1.25%	1.25	12,800	9.8%	1.0%	1.5%
India						
Women and men separately	0.92%	1.25	62,693	5.2%	0.8%	1.0%
Women and men together	0.92%	1.25	125,385	3.7%	0.9%	1.0%

Based on the assumed levels of HIV prevalence, the maximum permissible level of the relative standard errors and the design effects, it was determined that in each of the six high HIV prevalence states, 6,400 women and 6,400 men would be tested for HIV, and at the national level, the number tested would be about 125,000 women and men in about equal numbers. This sample size was sufficiently large to permit national estimates for urban and rural areas and for youth age 15-24 with a somewhat lower, but still acceptable, level of precision.

It was thus originally planned to have a total of about 76,800 persons tested in the six high HIV prevalence states. The balance of HIV tests needed to achieve a national sample of 125,385 HIV tests were to be distributed to all the remaining states in proportion to the sample size for individual interviews of women and men in those states. However, after the planning process had begun, it was decided to conduct 24,400 HIV tests in Uttar Pradesh in order to provide a separate estimate of HIV prevalence for this state. In light of this change, the final HIV sample design increased the required number of HIV tests nationally to 134,000, with approximately 101,000 tests in the seven states for which separate estimates were to be provided and 33,000 tests distributed between the remaining 22 states in proportion to their sample sizes.

Sample size for women's interviews

As many of the key indicators that were to be generated from NFHS-3, such as those pertaining to family planning practices and maternal and child health, refer to ever-married women in the reproductive ages, the target sample size in NFHS-3 was fixed in terms of ever-married women age 15-49 years. Based on previous survey experience, it was decided to select an initial target sample size of 1,500 completed interviews with ever-married women in states with a 2001 Census population less than 5 million, 3,000 completed interviews with ever-married women in states with a 2001 Census population between 5 and 30 million, and 4,000

completed interviews with ever-married women in states with a population of more than 30 million. The initial target sample size in Uttar Pradesh, the largest Indian state with about one-sixth of the country's population, was fixed at 10,000 ever-married women. The initial target sample size for the 29 states is given in the following table.

State	Initial target sample size for ever-married women age 15-49
Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura	1,500
Assam, Chhattisgarh, Delhi, Goa, Haryana, Himachal Pradesh, Jammu & Kashmir, Jharkhand, Punjab, Uttaranchal	3,000
Andhra Pradesh, Bihar, Gujarat, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Rajasthan, Tamil Nadu, West Bengal	4,000
Uttar Pradesh	10,000

However, in the states with the cities for which slum/non-slum estimates were to be produced and in the states for which separate estimates of HIV were to be produced, these initial state sample sizes were not sufficient and needed to be adjusted upwards.

Adjustments to the initial sample size in specific states. Based on the general assumption that a robust estimation of most population and health indicators requires a minimum sample of 1,000 ever-married women, the target sample size for the eight cities with separate slum and non-slum estimates was fixed at a minimum of 2,000 interviews with ever-married women (equally distributed between slum and non-slum areas of each city). In addition, adjustments needed to be made to the initial sample size in states with separate HIV prevalence estimates. These considerations resulted in the following adjustments to the initial sample size of ever-married women age 15-49 in specified states:

State	Adjusted target sample size of ever-married women 15-49	Reason for adjustment
Andhra Pradesh	6,500	To allow for a state-level estimate of HIV prevalence and slum/non-slum estimates for Hyderabad.
Karnataka	5,600	To allow for a state-level estimate of HIV prevalence.
Madhya Pradesh	6,000	To allow for slum/non-slum estimates for Indore.
Maharashtra	8,000	To allow for a state-level estimate of HIV prevalence and slum/non-slum estimates for Mumbai and Nagpur.
Manipur	4,250	To allow for a state-level estimate of HIV prevalence.
Nagaland	4,500	To allow for a state-level estimate of HIV prevalence.
Tamil Nadu	6,250	To allow for a state-level estimate of HIV prevalence and slum/non-slum estimates for Chennai.
Uttar Pradesh	11,000	To allow for slum/non-slum estimates for Meerut. (The initial target sample size was already sufficient to produce HIV estimates.)
West Bengal	5,750	To allow for slum/non slum estimates for Kolkata.

To the extent possible, each state sample size was allocated to urban and rural areas in proportion to the population in the 2001 Census. However, in all those states for which city-level estimates were required, a larger proportion of the target sample was allocated to urban areas.

The target sample size, which was fixed in terms of ever-married women, was transformed into the number of households to be covered by assuming the number of ever-married women per household was the same as observed in NFHS-2. With the assumption that on average 30 households would be interviewed in each primary sampling unit (PSU), the number of PSUs to be surveyed in urban and rural areas of all states and the eight cities selected for slum/non slum measures was estimated.

Sample size for men's interviews

It has already been mentioned that most of the population and health indicators to be measured in NFHS-3 refer to ever-married women and their children and are estimated from women's interviews. Estimation of these indicators, particularly for children, needs a fairly large sample size to produce robust estimates for subgroups of the population. The required sample size for men's interviews is considerably smaller because the men's interviews were not designed to produce indicator estimates for their children. Hence, it was decided to interview men only from a subsample of households except in the states for which separate HIV prevalence estimates were required. In those seven states, men from all the selected households would be interviewed.

Sampling Design

NFHS-3 adopted a two-stage sample design in most rural areas and a three-stage sample design in most urban areas. In each state, the rural sample was typically selected in two stages: the first stage involved selection of Primary Sampling Units (PSUs), i.e., villages, with probability proportional to population size (PPS); the second stage involved the systematic selection of households within each PSU.

The basic reason for adopting a three-stage sample design in urban areas is that urban wards are quite large, making it difficult to list all the households in a ward and select households directly from the resulting list. Although data on census enumeration blocks (CEBs) in all wards were collected in the 2001 Census, these data were not available in published form; hence, it was not possible to select CEBs directly from an existing list of CEBs for the urban part of a state. For this reason, the selection of the urban sample (in all urban areas except in the eight cities for which slum and non-slum estimates were to be provided) involved the following three stages: in the first stage, wards were selected with PPS sampling; in the next stage, one CEB was selected by PPS from each selected ward; and in the final stage, households were randomly selected within each selected CEB. However, CEB-wise data was made available by the Registrar General of India on special request for the eight cities for which slum/non-slum estimates were planned; hence in these eight cities, the sample households were selected in two stages. First slum and non-slum CEBs were selected, and then households were selected from selected CEBs.

Sample selection in rural areas

Sampling frame. In rural areas, the 2001 Census list of villages served as the sampling frame. The Primary Census Abstract (PCA) by the Census of India, 2001, provides data on the number

of households, persons, males, females, literates, persons belonging to the scheduled castes (SC), persons belonging to the scheduled tribes (ST), employed persons, etc. for all the villages in each district of India. All the villages with fewer than five households were removed from this list. Since the population in all such villages was only 0.02 percent of the total rural population, the exclusion of small villages from the sampling frame is not expected to disturb the representativeness of the sample. Further, to ensure the selection of a sufficient number of households in the selected PSUs, all the smaller villages with 5-49 households were linked with a nearby village. In some cases, more than two villages had to be linked to form a group of villages with a minimum of 50 households. In this way, a minimum size of 50 households in each village was ensured in all the villages in the sampling frame.

Stratification. To ensure the inclusion of villages with different socioeconomic characteristics in the sample, stratified sampling was adopted at the first stage of sample selection. All the villages (or groups of villages) were stratified by a number of variables. The first level of stratification was geographic, with districts being subdivided into contiguous regions. The categorization of districts in different regions in each state is shown in Table C.1. In some states, a single district comprised a region.

Within each of these regions, villages were further stratified using selected variables from the following list: village size, percentage of males working in the nonagricultural sector, percentage of the population belonging to scheduled castes or scheduled tribes, and female literacy. In addition to these variables, districts in high HIV prevalence states were stratified according to HIV prevalence. For this purpose, the classification of all the districts in the six high HIV prevalence states as 'high', 'medium', and 'low' prevalence districts by NACO was adopted. Female literacy was used for implicit stratification (i.e., the villages were ordered prior to selection according to the proportion of females who were literate) in most states, although female literacy was an explicit stratification variable in a few states. The list of variables used for explicit and implicit stratification in each state is given in Table C.2.

The number of PSUs to be selected from each stratum was proportional to the stratum size. In each stratum, an independent sample of PSUs was drawn using PPS sampling.

Sampling frame at the second stage. A household listing operation carried out in each sample area provided the necessary frame for selecting households at the second stage. The household listing operation involved preparing up-to-date location and layout sketch maps, assigning a number to each structure, recording addresses of these structures, identifying residential structures, and then listing the names of heads of households. Listing of all households in large villages with 500 or more households is a huge task, potentially prone to errors of omission or duplication. Hence, large villages with about 500 or more households were segmented into three or more segments (depending on village size) of approximately equal size (usually about 100 to 200 households). From all the segments in each large village, two segments were selected using PPS sampling. Houselisting was then done only in the two selected segments. In all such large villages, the sampling design became a three-stage design.

The selection of the required number of households was done using systematic sampling. For the logistical convenience of the fieldwork and to minimize refusals, runs of five consecutive households, rather than single households, were selected systematically.

Selection of households. The number of households selected in a rural PSU was the product of number of households listed in the PSU and the probability of selection of a household in the selected rural PSU.

The probability of selecting a household from a selected rural PSU (f_2) of a specific state was computed as:

$$f_2 = \frac{f}{f_1}$$

where f, the domain sampling fraction (i.e., the probability of selecting an ever-married woman in the rural area of a state), is computed as:

$$f = \frac{n}{N}$$

where

n = number of ever-married rural women to be interviewed (after adjusting upward to account for nonresponse and other loss), and

N = projected rural population of ever-married women in the state in March 2006 (the midpoint of the NFHS-3 survey period).

The probability of selecting a PSU from a rural stratum of a state (f_1) was computed as:

$$f_1 = \frac{a * s_i}{\sum s_i}$$

where

a = number of rural PSUs selected from the stratum,

s_i = population size of the i^{th} PSU, and

Σs_i = total rural population of the stratum.

Sample selection in urban areas

The following discussion refers to sampling in all the urban areas of the states in which no separate city estimates were planned. In Andhra Pradesh, Madhya Pradesh, Maharashtra, Uttar Pradesh, Tamil Nadu, and West Bengal, the following procedure was adopted for sampling in all urban areas excluding the cities in which slum/non-slum estimates were planned. This procedure was not applied to sampling in urban areas of Delhi.

Sampling frame. In urban areas, the sampling frame for the first stage of selection was the list of all the wards in a state. The Primary Census Abstract (PCA) provides details of all the wards in each city or town in urban areas. The list was arranged by geographical regions and then by female literacy. A sample of urban wards was drawn from the list of wards by PPS sampling.

Sample selection at the second and third stages. As mentioned earlier, the sample of households from a selected ward was drawn by adopting two-stage sampling. A ward in a

township is a large area comprising a large number of households. Each ward comprises several enumeration blocks (CEB) created for the census. A list of all the CEBs in a selected ward formed the sampling frame at the second stage. Such lists of CEBs in the selected wards were made available for use for NFHS-3 by the census office on request. Each CEB is comprised of about 150-200 households. A sample of households from a ward was drawn in two stages. One CEB was selected from each selected ward at the first stage. Then in each selected CEB, a household listing operation was carried out (similar to the listing operation in rural PSUs). The household listing provided the necessary frame for selecting households at the third stage. The selection of the required number of households from the list of households was done using systematic sampling. As in rural areas, systematically selected groups of five consecutive households, rather than a single household, were selected.

Similar to rural PSUs, the sample size of households was not the same in every urban PSU. The sample size of households in an urban PSU was obtained by applying the probability of selection of an ever-married woman in an urban PSU (f_3) to the total number of households listed in the urban PSU.

The domain sampling fraction, i.e., the probability of selecting an ever-married woman from an urban domain in a state (f), was computed as:

$$f = \frac{n}{N}$$

where

n = number of ever-married women to be interviewed from the urban domain (after adjusting upward to account for nonresponse and other loss), and

N = projected population of ever-married women in the urban domain in March 2006.

The probability of selecting a ward from an urban stratum (f_1) was computed as:

$$f_1 = \frac{a * s_i}{\sum s_i}$$

where

a = number of wards selected from the stratum,

s_i = population size of the i^{th} ward, and

Σs_i = total population of the urban stratum.

The probability of selecting a CEB from a selected ward (f_2) was computed as:

$$f_2 = \frac{B_i}{\sum B_i}$$

where

B_i = population size of the i^{th} CEB, and

ΣB_i = total population of the ward.

The probability of selecting a household from a selected block (f_3) was computed as:

$$f_3 = \frac{f}{f_1 * f_2}$$

Sample selection in eight cities

In the eight cities for which separate indicators for slum and non-slum areas were to be provided, a different sampling procedure was adopted. As mentioned above, sample households in the cities were selected with a two-stage sampling design. At the first stage CEBs were selected and at the second stage households were selected. In each city, slum and non-slum PSUs were selected independently from the respective lists of slum and non-slum CEBs. For the estimation of the number of households and the selection of households from the selected slum/non-slum PSUs (CEBs), a procedure similar to that used for rural and urban PSUs was adopted.

Sampling frame. As mentioned above, although the Census of India collects and compiles data by CEBs, that information is not published. However, on special request for NFHS-3, the Registrar General of India made available CEB-wise data for the eight cities for which slum and non-slum estimates were to be provided. Two separate lists of all the slum CEBs and non-slum CEBs in all the wards of each city served as two separate sampling frames at the first stage of selection. From each sampling frame, slum and non-slum PSUs were selected using PPS sampling. The houselisting carried out in each of the selected CEBs served as the sampling frame for the selection of households.

Sample weights

The basic objective of weighting sample data is to try and maximize the representativeness of the sample in terms of the size, distribution, and characteristics of the study population. When sample units have been selected with differing probabilities, it is common to weight the results inversely proportional to the unit selection probabilities, i.e., the design weight, so as to reflect the actual situation in the population. In a survey sample selected from a robust frame and well implemented with high response rates, the application of the design weight is all that is required. In practice, however, the situation is more complicated because of shortcomings in the selection and implementation of the sample.

In NFHS-3, two sets of weights are in operation. One set of weights is used for generating national-level indicators and another set is used for producing state-level indicators. Each set has the following different types of weights:

1. *Household weight* for estimating indicators such as the proportion of female-headed households and the percentage of children age 0-4 years whose birth was registered.
2. *Woman's weight* for generating indicators such as the contraceptive prevalence rate and the percentage of children age 12-23 months who are fully immunized.
3. *Man's weight* for estimation of indicators such as the proportion of men who have heard of AIDS.

4. *HIV weight* for women for generating indicators such as HIV prevalence by age for women.
5. *HIV weight* for men for generating indicators such as HIV prevalence by age for men.
6. *Domestic violence weight* for estimating indicators such as the proportion of married women who have ever experienced spousal violence.

Note that children of interviewed women are assigned the weight of their mother. Some child indicators are assessed at the household level; in such cases, children are assigned the weight of the household.

Calculation of sampling weights

Calculation of state and national household weights. The basic reasons for weighting primary data while estimating state-level indicators are:

a) To take care of the non-equal probability of selection in different domains, i.e., rural and urban areas and slum and non-slum areas in the seven states of Andhra Pradesh, Delhi, Madhya Pradesh, Maharashtra, Tamil Nadu, Uttar Pradesh, and West Bengal. In each state, the total sample size was distributed between urban and rural areas according to the proportion of urban-rural households. However, oversampling was done in urban areas in states with very small urban populations in order to have a sample large enough to yield stable estimates; oversampling was also done in the cities where slum and non-slum estimates were required in order to have large enough samples in the slum areas and the non-slum areas separately. Whatever the reason, oversampling of urban areas of some states leads to unequal probabilities of selection.

b) To take care of the differential non-response rates of household interviews in different domains, namely urban and rural areas and slum and non-slum areas of the eight cities.

To take care of the non-equal probabilities of selection in different domains, a design weight was computed. The household design weight W_{Di} for the i^{th} domain is the inverse of the sampling fraction for the i^{th} domain ($f_i = n_i/N_i$). To take care of differential nonresponse in different domains, the design weight for each domain is multiplied by the inverse of the response rates.

The household weight (W_{Hi}) for the i^{th} domain is then:

$$W_{Hi} = \frac{W_{Di}}{R_{Hi}}$$

where

R_{Hi} = response rate of the household interview (number of completed household interviews divided by the number of households selected for interview)

After adjustment for nonresponse, the weights are normalized so that the total number of weighted cases is equal to the total number of unweighted cases. This is done by multiplying W_{Hi} for each domain by the ratio of total number of unweighted cases to total number of weighted cases (obtained by applying weights before normalization to the number of cases in each domain).

The final state household weight is calculated as

$$\hat{W}_{Hi} = W_{Hi} * \left[\frac{\sum n_i}{\left(\sum W_{Hi} * n_i\right)} \right]$$

Because of the normalization of the state household weight at the state level, the normalized state household weight cannot be used for national indicators. A set of national weights is thus calculated. The final national household weight is based on the state household weight W_{Hi} (after correction for nonresponse and before normalization at the state level) normalized at the national level.

Calculation of state and national women's and men's weights. The state level and national level women's weights are derived exactly the same way as the household weights. The only difference is that the response rate of the household interview is replaced by the response rate of the household interview (R_{Hi}) times the response rate of the women's interview (R_{Wi}). The eligible women's weight is:

$$W_{Wi} = \frac{W_{Di}}{R_{Hi} * R_{Wi}}$$

Normalization of W_{Wi} at the individual state level gives the state's normalized women's weight; normalization at the national level for all states together gives the all-India normalized women's weight.

Men's weights are calculated in the same way by incorporating the men's response rates.

The state and national HIV household weights and HIV individual weights are also calculated in a similar fashion. The only difference is the normalization of the individual HIV weights. Since HIV prevalence is reported not only for women and men separately, but also for women and men together, the individual HIV weights must be normalized for women and men together at the individual state level and at the national level, respectively.

Calculation of state and national domestic violence weights. Unlike all other sections of the Woman's Questionnaire, the section on domestic violence was not administered to all women in the household. This was part of a strategy built into the questionnaire to maintain confidentiality and to protect respondents from any potential security-related problems. Only one eligible woman in each selected household was administered the section on domestic violence. In households with more than one eligible woman, the woman that the module was administered to was randomly selected through a specially designed simple selection procedure based on the

Kish Grid (see the NFHS-3 Household Questionnaire) which was built into the Household Questionnaire.

The selection of one woman per sample household implies that women who were selected for the domestic violence module of the questionnaire are a subsample of the entire NFHS-3 sample. This subsampling results in the need to generate one additional set of weights for estimating state and national indicators of domestic violence. The domestic violence weights (DW_{Wi}) are calculated the same way as other weights but they also take account of the number of eligible women in the household.

The domestic violence weight for women is:

$$DW_{Wi} = \frac{W_{Di}}{R_{Hi} * S_{Hi} * DR_{Wi}}$$

where

S_{Hi} = the inverse of the number of eligible women in the household, and
DR_{wi} = response rate of the women selected for the domestic violence module.

Normalization of DW_{Wi} at the individual state level gives the state's normalized domestic violence weight; normalization at the national level for all states together gives the all-India normalized domestic violence weight.

Sample implementation

NFHS-3 response rates for households, women, and men were discussed in Chapter 1, Section 1.7. Tables C.3.1 and C.3.2 provide additional information on the reasons for non-response for women and men, respectively. Response rates for HIV blood collection were estimated separately and were discussed in Chapter 12.

Table C.1 Regions by state

State	Region	2001 Census districts
Andhra Pradesh	I	Srikakulam, Vizianagaram, Visakhapatnam
	II	East Godavari, West Godavari, Krishna, Guntur
	III	Prakasam, Nellore
	IV	Chittoor, Cuddapah, Anantapur, Kurnool
	V	Mahbubnagar, Rangareddi, Hyderabad, Medak, Nizamabad, Adilabad, Nalgonda
	VI	Karimnagar, Warangal, Khammam
Arunachal Pradesh	I	Tawang, West Kameng, East Kameng
	II	Papumpare, Lower Subansiri, Upper Subansiri
	III	West Siang, East Siang, Upper Siang
	IV	Darrang, Lohit
	V	Tirap, Changlang
Assam	I	Goalpara, Kamrup, Marigaon, Nagaon, Darrang
	II	Dhubri, Bongaigaon, Barpeta, Nalbari, Sonitpur
	III	Golaghat, Jorhat, Sibsagar, Dibrugarh, Tinsukia, Karimganj, Hailakandi, Cachar
	IV	Lakhimpur, Dhemaji
	V	Karbi Anglong
Bihar	I	Saran, Siwan, Gopalganj, Pashchim Champaran, Purba Champaran
	II	Sitamarhi, Vaishali, Darbhanga, Madhubani, Samastipur, Muzaffarpur, Saharsa, Sheohar, Supaul
	III	Purnia, Araria, Kishanganj, Katihar, Madhepura
	IV	Patna, Bhojpur, Rohtas, Buxor, Kaimur (Bhabua)
	V	Munger, Khagaria, Bhagalpur, Banka, Jamui, Lakhisarai, Sheikhpura
	VI	Nalanda, Gaya, Jehanabad, Aurangabad, Nawada, Begusarai
Chhattisgarh	I	Koriya, Surguja, Bilaspur, Korba, Janjgir-champa, Jashpur, Raigarh
	II	Kawaedha, rajnandgao, Durg, Raipur, Mahasamund, Dhamtari
	III	Kanker, Bastar, Dantewada
Goa		Each district is one region
Gujarat	I	Jamnagar, Rajkot
	II	Surendranagar, Bhavnagar, Amreli
	III	Junagadh, Porbandar
	IV	Kachchh, Banaskantha
	V	Sabarkantha, Mahesana, Patan
	VI	Gandhinagar, Ahmedabad
	VII	Kheda, Anand
	VIII	Panch Mahals, Vadodara, Dohad
	IX	Bharuch, Surat, Valsad, The Dangs, Narmada, Navsari
Haryana	I	Faridabad, Gurgaon, Sonipat
	II	Ambala, Kaithal, Karnal, Kurukshetra, Panipat, Yamunanagar, Panchkula
	III	Bhiwani, Mahendragarh, Rewari, Sirsa
	IV	Hissar, Jind, Rohtak, Fatehbad, Jhajjar
Himachal Pradesh	I	Chamba, Lahaul-Spiti, Kinnaur
	II	Kangra, Hamirpur, Una, Mandi, Kullu
	III	Bilaspur, Shimla, Solan, Sirmaur
Jammu & Kashmir	I	Kargil, Leh
	II	Anantnag, Pulwama, Srinagar, Badgam, Baramulla
	III	Kupwara, Rajouri, Poonch
	IV	Doda, Udhampur, Kathua, Jammu
Jharkhand	I	Deogarh, Godda, sahibganj, Pakaur, Dumka
	II	Chatra, Hazaribagh, Kodarma, Giridh, Bokaro, Dhanbad
	III	Garhwa, Palamu, Lohardagga, Gumla, Ranchi, Pashchim Singhbhum, Purbi Singhbhum
Karnataka	I	Bidar, Bijapur, Gulbarga, Raichur, Bagalkot, Koppal
	II	Belgaum, Dharwad, Gadag, Haveri
	III	Dakshina Kannada, Kodagu, Uttara Kannada, Udupi
	IV	Chikmagalur, Shimoga
	V	Bangalore rural, Bellary, Chitradurga, Kolar, Tumkur, Davangere
	VI	Hassan, Mandya, Mysore, Chamrajnagar
Kerala		Each district is one region

Continued…

Table C.1 Regions by state—*Continued*

State	Region	2001 Census districts
Madhya Pradesh	I	Panna, Rewa, Satna, Sidhi, Shahdol, Chhatarpur, Tikamgarh, Umaria
	II	Raisen, Sagar, Damoh, Vidisha, Bhopal, Sehore
	III	Dewas, Dhar, Indore, Jhabua, Ujjain, Rajgarh, Ratlam, Mandsaur, Shajapur, Neemuch
	IV	Mandla, Jabalpur, Seoni, Narsimhapur, Chhindwara, Balaghat, Dindori, Katni
	V	Betul, Hoshangabad, East Nimar, West Nimar, Barwani, Harda
	VI	Gwalior, Bhind, Morena, Datia, Guna, Shivpuri, Sheopur
Maharashtra	I	Thane, Raigad, Ratnagiri, Sindhudurg
	II	Nasik, Dhule, Jalgaon, Nandurbar
	III	Ahmednagar, Pune, Satara, Sangli, Solapur, Kolhapur
	IV	Aurangabad, Jalna, Parbhani, Bid, Latur, Osmanabad, Buldhana, Akola, Amaravati, Washim, Hingoli
	V	Yeotmal, Wardha, Nagpur, Nanded
	VI	Bhandara, Chandrapur, Gadchiroli, Gondiya
Manipur	I	Senapati, Tamenglang, Ukhrul
	II	Churchandpur, Chandel
	III	Bishnupur , Imphal West, Imphal East, Toubal
Meghalaya	I	West Garo Hills
	II	East Garo Hills, South Garo Hills
	III	West Khasi Hills
	IV	Ri Bhoi, East Khasi Hills
	V	Jantia Hills
Mizoram	I	Kolasib, Mamit, Aizwal, Champhal, Serchip
	II	Lunglei
	III	Lawangttai, Saita
Nagaland	I	Tuensang, Mon
	II	Mokakchung, Zunhebota
	III	Wokha , Dimapur
	IV	Kohima , Phek
Orissa	I	Sundargarh, Kendujhar, Mayurbhanj
	II	Phulabani, Koraput, Baudh, Malkangiri, Navarangapur, Rayagada
	III	Sambalpur, Balangir, Kalahandi, Bargarh, Debagarh, Jharsuguda, Nuapada, Sonapur
	IV	Baleshwar, Cuttack, Ganjam, Puri, Dhenkanal, Anugul, Bhadrak, Gajapati, Jagatsinghpur, Jajapur, Kendrapara, Khordha, Nayagarh
Punjab	I	Gurdaspur, Amritsar, Firozpur
	II	Jalandhar, Kapurthala, Hoshiarpur, Rupnagar, Nawashahr
	III	Ludhiana, Patiala, Sangrur, Fatehgarh Sahib
	IV	Bhatinda, Faridkot, Mansa, Moga, Muktsar
Rajasthan	I	Ganganagar, Bikaner, Churu, Jaisalmer, Jodhpur, Nagaur, Pali, Barmer, Jalor, Sirohi, Hanuman garh
	II	Jhunjhunun, Alwar, Bharatpur, Dhaulpur, Sawai Madhopur, Jaipur, Sikar, Ajmer, Tonk, Bhilwara, Dausa, Karauli
	III	Dungarpur, Banswara, Udaipur, Rajasamand
	IV	Chhitaurgarh, Bundi, Kota, Jhalawar, Baran
Sikkim		Each district is one region
Tripura		Each district is one region
Tamil Nadu	I	Coimbatore, Dindigul Anna (Dindigul), Madurai (Madura and Theni), Periyar (Erode), Nilgiri
	II	North Arcot-Ambedkar (Vellore), Dharmapuri, Tiruvannamalai-Sambuvarayan, Salem (Salem and Namakal), Tiruchirappalli (Tiruchirapalli, Karur and Perambalur), Ariyalur
	III	Kanniyakumari
	IV	Chengalpattu-MGR (Kanchipuram and Tiruvallur), South Arcot (Cuddalore and Villupuram), Thanjavur, (Thanjavur, Nagappattinam and Tiruvarur), Chennal
	V	Pudukkottai, Pasumpon, Muthuramalinga Thevar (Sivagangai), Kamarajar (Virudhunagar), Ramanathapuram, Chidambaranar (Thootukudi), Tirunelveli Kattabomman (Tirunelveli)
Uttaranchal	I	Dehradun, Hardwar, Udham Singh Nagar
	II	Uttar Kashi, Tehri Garhwal, Rudra Prayag, Pauri Garhwal
	III	Chamoli, Bageshwar, Pitorgarh, Almora, Nainital, Champawat

Continued…

Table C.1 Regions by state—*Continued*

State	Region	2001 Census districts
Uttar Pradesh	I	Bijnor, Ghaziabad, Meerut, Moradabad, Rampur, Saharanpur, Muzzafarnagar, Agra, Aligarh, Bareilly, Budaun, Bulandshahr, Etah, Farrukhabad, Firozabad, Mainpuri, Pilibhit, Shahjahanpur, Etawah, Mathura, Auraiya, Baghpat, Goutam Buddha Nagar, Hathras, Jyotiba Phule Nagar, Kannauj
	II	Kheri, Hardoi, Rae Bareli, Sitapur, Barabanki, Fatehpur, Kanpur Dehat, Kanpur Nagar, Lucknow, Unnao
	III	Allahabad, Gonda, Pratapgarh, Sultanpur, Bahraich, Faizabad, Azamgarh, Basti, Deoria, Gorakhpur, Jaunpur, Maharajganj, Mau, Siddharthnagar, Ballia, Gazipur, Varanasi, Mirzapur, Sonbhadra, Ambedkar Nagar, Balrampur, Chandauli, Kaushambi, Kushinagar, Sant Kabir Nagar, Sant Ravidas Nagar, Shrawasti
	IV	Banda, Lalitpur, Hamirpur, Jalaun, Jhansi, Chitrakoot, Mahoba
West Bengal	I	Jalpaiguri, Darjeeling
	II	Koch Bihar, West Dinajpur (Uttar Dinajpur and Dakshin Dinajpur), Maldah, Murshidabad
	III	Nadia, Haora, Hugli, North Twenty-Four Parganas, South Twenty-Four Parganas, Barddhaman
	IV	Medinipur, Bankura, Birbhum
	V	Puruliya

Note: Delhi does not have regions.

Table C.2 Sample characteristics

State	2001 population[1]	Variables for rural stratification	Number of primary sampling units				
			Urban			Rural	Total
			Slum	Non-slum	Total		
Andhra Pradesh	76,208,158	Region (group of districts) HIV Prevalence Village size Percent of SC/ST population Female literacy (implicit)	28	31	91	104	195
Arunachal Pradesh	1,093,199	Region (group of districts) Village size Female literacy (implicit)	na	na	17	48	65
Assam	26,651,050	Region (group of districts) Village size Percent of SC/ST population Percent of males in nonagricultural work Female literacy (implicit)	na	na	29	58	87
Bihar	82,991,525	Region (group of districts) Village size Percent of SC/ST population Female literacy (implicit)	na	na	31	71	102
Chhattisgarh	20,832,242	Region (group of districts) Village size Female literacy Percent of SC/ST population (implicit)	na	na	30	64	94
Delhi	13,850,477	Village size Female literacy (implicit)	37	67	104	8	112
Goa	1,347,641	Region Village size Female literacy (implicit)	na	na	63	63	126
Gujarat	50,670,013	Region (group of districts) Village size Percent of SC/ST population Female literacy (implicit)	na	na	43	70	113
Haryana	21,144,059	Region (group of districts) Village size Percent of SC/ST population Female literacy (implicit)	na	na	29	62	91
Himachal Pradesh	6,064,355	Region (group of districts) Village size Percent of SC/ST population Female literacy (implicit)	na	na	34	72	106
Jammu & Kashmir	10,142,775	Region (group of districts) Village size Percent of SC/ST population Percent of males in nonagricultural work Female literacy (implicit)	na	na	32	65	97
Jharkhand	26,937,511	Region (group of districts) Village size Female literacy Percent of SC/ST population (implicit)	na	na	34	61	95
Karnataka	52,843,804	Region (group of districts) HIV Prevalence Village size Percent of SC/ST population Female literacy (implicit)	na	na	64	112	176
Kerala	31,841,187	Region (group of districts) Female literacy	na	na	34	91	125
Madhya Pradesh	60,336,893	Region (group of districts) Village size Percent of SC/ST population Female literacy (implicit)	30	33	94	92	186

Continued…

Table C.2 Sample characteristics—*Continued*

State	2001 population[1]	Variables for rural stratification	Number of primary sampling units				
			Urban			Rural	Total
			Slum	Non-slum	Total		
Maharashtra	96,874,368	Region (group of districts) HIV Prevalence Village size Percent of SC/ST population Percent of males in nonagricultural work Female literacy (implicit)	72	80	203	86	289
Manipur	2,166,589	Region (group of districts) HIV Prevalence Village size Female literacy (implicit)	na	na	57	98	155
Meghalaya	2,317,408	Region (group of districts) Village size Percent males in nonagricultural work Female literacy (implicit)	na	na	25	46	71
Mizoram	888,451	Region (group of districts) Village size Percent of SC/ST population (implicit)	na	na	31	33	64
Nagaland	1,990,021	Region (group of districts) HIV Prevalence Village size Female literacy (implicit)	na	na	76	125	201
Orissa	36,792,823	Region (group of districts) Village size Female literacy Percent of males in nonagricultural work Percent of SC/ST population (implicit)	na	na	32	83	115
Punjab	24,356,831	Region (group of districts) Village size Percent of SC/ST population Percent of males in nonagricultural work Female literacy (implicit)	na	na	34	65	99
Rajasthan	56,500,901	Region (group of districts) Village size Percent of SC/ST population Percent of males in nonagricultural work Female literacy (implicit)	na	na	32	74	106
Sikkim	540,768	Region Village size Percent of SC/ST population Female literacy (implicit)	na	na	19	41	60
Tamil Nadu	62,405,273	Region (group of districts) HIV Prevalence Village size Percent of SC/ST population Female literacy (implicit)	32	36	110	104	214
Tripura	3,199,145	Region Village size Percent of SC/ST population Female literacy (implicit)	na	na	17	39	56
Uttaranchal	8,480,537	Region (group of districts) Village size Percent of SC/ST population Female literacy (implicit)	na	na	32	66	98
Uttar Pradesh	166,171,011	Region (group of districts) Village size Percent of SC/ST population Percent of males in nonagricultural work Female literacy (implicit)	33	36	133	220	353

Continued…

Table C.2 Sample characteristics—*Continued*

State	2001 population[1]	Variables for rural stratification	Number of primary sampling units				
			Urban			Rural	Total
			Slum	Non-slum	Total		
West Bengal	80,171,814	Region (group of districts) Village size Percent of SC/ST population Percent of males in non-agricultural work Female literacy (implicit)	39	43	119	86	205
India	1,025,810,829		271	326	1,649	2,201	3,850

na = Not applicable
SC = Scheduled caste
ST = Scheduled tribe
[1] The population shown is the 2001 Census population, excluding persons living in villages with fewer that five households.

Table C.3.1 Sample implementation: Women

Percent distribution of households and eligible women by results of the household and individual interviews, and household, eligible women, and overall response rates, by residence, India, 2005-06

Result	Residence		Total
	Urban	Rural	
Selected households			
Completed (C)	92.3	94.5	93.5
Household present but no competent respondent at home (HP)	1.4	1.1	1.2
Postponed (P)	0.1	0.0	0.0
Refused (R)	1.4	0.2	0.8
Dwelling not found (DNF)	0.1	0.1	0.1
Household absent (HA)	2.2	2.0	2.1
Dwelling vacant/address not a dwelling (DV)	2.5	1.9	2.1
Dwelling destroy (DD)	0.1	0.2	0.1
Total	100.0	100.0	100.0
Number of sampled households	54,453	62,199	116,652
Household response rate (HRR)[1]	96.9	98.5	97.7
Eligible women			
Completed (EWC)	93.3	95.5	94.5
Not at home (EWNH)	3.4	2.6	2.9
Postponed (EWP)	0.1	0.0	0.1
Refused (EWR)	2.3	0.8	1.5
Partly completed (EWPC)	0.3	0.2	0.2
Incapacitated (EWI)	0.4	0.6	0.5
Other (EWO)	0.3	0.2	0.3
Total	100.0	100.0	100.0
Number of women	61,028	70,568	131,596
Eligible women response rate (EWRR)[2]	93.3	95.5	94.5
Overall response rate (ORR)[3]	90.4	94.1	92.4

[1] Using the number of households falling into specific response categories, the household response rate (HRR) is calculated as:

$$\frac{100 * C}{C + HP + P + R + DNF}$$

[2] Using the number of eligible women falling into specific response categories, the eligible women response rate (EWRR) is calculated as:

$$\frac{100 * EWC}{EWC + EWNH + EWP + EWR + EWPC + EWI + EWO}$$

[3] The overall response rate (ORR) is calculated as:

$$ORR = HRR * EWRR/100$$

Table C.3.2 Sample implementation: Men

Percent distribution of households and eligible men by results of the household and individual interviews, and household, eligible men, and overall response rates, by residence, India, 2005-06

Result	Residence		Total
	Urban	Rural	
Selected households			
Completed (C)	91.8	94.3	93.1
Household present but no competent respondent at home (HP)	1.3	1.2	1.3
Postponed (P)	0.1	0.0	0.1
Refused (R)	1.7	0.2	0.9
Dwelling not found (DNF)	0.1	0.1	0.1
Household absent (HA)	2.2	2.1	2.1
Dwelling vacant/address not a dwelling (DV)	2.6	1.9	2.2
Dwelling destroyed (DD)	0.2	0.2	0.2
Total	100.0	100.0	100.0
Number of sampled households	36,313	37,661	73,974
Household response rate (HRR)[1]	96.6	98.4	97.5
Eligible men			
Completed (EMC)	84.6	89.9	87.1
Not at home (EMNH)	11.2	7.6	9.5
Postponed (EMP)	0.1	0.0	0.1
Refused (EMR)	3.0	1.1	2.1
Partly completed (EMPC)	0.1	0.1	0.1
Incapacitated (EMI)	0.5	0.8	0.7
Other (EMO)	0.4	0.4	0.4
Total	100.0	100.0	100.0
Number of men	45,133	40,240	85,373
Eligible men response rate (EMRR)[2]	84.6	89.9	87.1
Overall response rate (ORR)[3]	81.8	88.4	84.9

[1] Using the number of households falling into specific response categories, the household response rate (HRR) is calculated as:

$$\frac{100 * C}{C + HP + P + R + DNF}$$

[2] Using the number of eligible men falling into specific response categories, the eligible men response rate (EMRR) is calculated as:

$$\frac{100 * EMC}{EMC + EMNH + EMP + EMR + EMPC + EMI + EMO}$$

[3] The overall response rate (ORR) is calculated as:

ORR = HRR * EMRR/100