

Maharashtra, Manipur, Tamil Nadu, and Uttar Pradesh. All women age 15-49 and men age 15-54 in the sample households were eligible for interviewing in all of these states plus Nagaland. In the remaining 22 states, all ever-married and never married women age 15-49 in sample households were eligible to be interviewed. In those 22 states, men age 15-54 were eligible to be interviewed in only a subsample of households. HIV tests for women and men were carried out in only a subsample of the households that were selected for men's interviews in those 22 states. The reason for this sample design is that the required number of HIV tests is determined by the need to calculate HIV prevalence at the national level and for some states, whereas the number of individual interviews is determined by the need to provide state level estimates for attitudinal and behavioural indicators in every state. For statistical reasons, it is not possible to estimate HIV prevalence in every state from NFHS-3 as the number of tests required for estimating HIV prevalence reliably in low HIV prevalence states would have been very large.

Sample Design



Strata

The urban and rural samples within each state were drawn separately and, to the extent possible, unless oversampling was required to permit separate estimates for urban slum and non-slum areas, the sample within each state was allocated proportionally to the size of the state's urban and rural populations. A uniform sample design was adopted in all states. In each state, the rural sample was selected in two stages, with the selection of Primary Sampling Units (PSUs), which are villages, with probability proportional to population size (PPS) at the first stage, followed by the random selection of households within each PSU in the second stage. In urban areas, a three-stage procedure was followed. In the first stage, wards were selected with PPS sampling. In the next stage, one census enumeration block (CEB) was randomly selected from each sample ward. In the final stage, households were randomly selected within each selected CEB.

Sample Selection in Rural Areas

In rural areas, the 2001 Census list of villages served as the sampling frame. The list was stratified by a number of variables. The first level of stratification was geographic, with districts being subdivided into contiguous regions. 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, an external estimate of HIV prevalence, i.e., 'High', 'Medium' or 'Low', as estimated for all the districts in high HIV prevalence states, was used for stratification in high HIV prevalence states. Female literacy was used for implicit stratification (i.e., villages were ordered prior to selection according to the proportion of females who were literate) in most states although literacy was an explicit stratification variable in a few states.

In every state, a mapping and household listing operation was carried out in each sample area. The listing provided the necessary frame for selecting households at the second stage. The household listing operation involved preparing up-to-date notional and layout sketch maps of each selected PSU, assigning numbers to structures, recording addresses or the location of these structures, identifying residential structures, and listing the names of the heads of all the households in residential structures in the selected PSUs. Large sample villages (with more than

a specified number of households, usually 500) were segmented, and two segments were selected randomly using the PPS method. Household listing in the segmented PSUs was carried out only in the selected segments. Each household listing team comprised one lister and one mapper. Senior field staff of the concerned research organization supervised the listing operation.

The households to be interviewed were selected with equal probability from the household list in each area using systematic sampling. The interval applied for the selection was determined to obtain a self-weighting sample of households within each domain. On average, 30 households were initially targeted for selection in each selected enumeration area. To avoid extreme variations in the workload, minimum and maximum limits were put on the number of households that could be selected from any area, at 15 and 60, respectively. Each survey team supervisor was provided with the original household listing, layout sketch map, and the list of selected households for each PSU. All the households which were selected were contacted during the main survey, and no replacement was made if a selected household was absent during data collection. However, if a PSU was inaccessible, a replacement PSU with similar characteristics was selected by IIPS and provided to the research organization.

Sample Selection in Urban Areas

The procedure adopted for the first stage of the sample design in urban areas was similar to the one followed in rural areas. The 2001 Census list of wards was arranged according to districts and within districts by the level of female literacy, and a sample of wards was selected systematically with probability proportional to size. Next, one census enumeration block, consisting of approximately 150-200 households, was selected from each selected ward using the PPS method. As in rural areas, a household listing operation was carried out in each selected census enumeration block, which provided the necessary frame for selecting households in the third stage of sample selection. On average, 30 households were targeted for selection from each census enumeration block with minimum and maximum limits from any area of 15 and 60 households.

Sample Weights

NFHS-3 is designed for self weighting at the domain level. The domains are the urban and rural areas of each state, and the slum and non-slum areas of each of the eight selected cities. This means that all households and individuals in the same domain will share a common household weight and individual weight, respectively. The design weight is the inverse of the overall sampling fraction in each domain. The overall sampling fraction is the product of the selection probabilities at each sampling stage (two stages in rural areas and three stages in urban areas). The design weight was adjusted for household non-response in the calculation of the household sampling weight. The household sampling weight was further adjusted for individual non-response to obtain the individual sampling weight. Both adjustments for non-response were done at the domain level in order to preserve the self-weighting nature of the sample within domains. The sampling weights were further normalized at the national level to obtain national standard weights and at the state level to obtain standard state weights for each of the 29 states. The national standard weights were normalized so that the total number of weighted cases equals the total number of unweighted cases at the national level. The state standard weights were calculated to ensure that the total number of weighted cases equals the total number of