C. Annexure 03 – Sample weights

First, appropriate sampling weights for households were constructed for each state data set separately for Urban and Rural areas. The element weight consisted of factors reflecting ward selection probabilities, Census Enumeration Block (CEB) selection probabilities within wards, household selection probabilities within CEB and household non-response adjustments.

In urban areas, from the list of wards, 300 wards were selected with PPS and from each ward 20 household were selected using circular systematic sampling.

For urban areas, the weight HWT_{ijk} for the household k in CEB j of ward i, can be expressed as follows

$HWT_{ijk} = W_{1i} * W_{2j|i} * W_{3k|i,j}$ i=1,2,.....,300 , j=1, k=1,2,.....,20

where $W_{1i} = \frac{1}{\pi_i}$: the reciprocal of the inclusion probability π_i of ward i

where $\pi_i = \frac{a \times \text{Population of ward i}}{\text{Total urban population}}$ and a (=300) was the total number of wards to be selected from the urban areas.

$W_{2j/i} = \frac{1}{\pi_{j|i}}$: the reciprocal of the conditional probability of selection of CEB j in ward i

where $\pi_{j|i} = \frac{\text{Population of selected CEB j within ward i}}{\text{Population of selected ward i}}$

$W_{3k|i,j} = \frac{1}{\pi_{k|i,j} \times \hat{\theta}_{k|i,j}}$: the reciprocal of the product of conditional inclusion probability $\pi_{k|i,j}$ of household k in the j[th] selected CEB of the i[th] ward and estimated conditional response probability $\hat{\theta}_{k|i,j}$ of household k from within the j[th] selected CEB of ward i.

where $\pi_{k|i,j} = \frac{\text{Number of households sampled from selected CEB j of ward i}}{\text{Number of households in selected CEB j of ward i}}$

$$HWT_{ijk} = \frac{\text{Size of urban population}}{20 \times \text{Population of selected CEB from ward i}} \times \frac{\text{Number of households in selected CEB of ward i}}{\text{Number of households sampled from selected CEB of ward i with HH result code completed}}$$

In rural areas, from the lists of villages, 300 villages (or cluster of villages) were selected with probability proportional to size and from each village 20 households were selected using systematic sampling.

Proceeding as above it can be shown that the weight for the k^{th} selected household of the i^{th} selected village, HWT_{ik},

$$HWT_{ik} = \frac{\text{Size of rural population}}{20 \times \text{Population of } i^{th} \text{ selected village}} \times$$

$$\frac{\text{Number of households in } i^{th} \text{ selected village}}{\text{Number of households selected from } i^{th} \text{ village with HH result code complete}}$$

INDIVIDUAL WEIGHTS

From each selected household one member aged 18-69 were selected using the KISH method. Post stratification weights for individuals were constructed using the state age distributions for both gender of the urban areas which are available on the population level. At first, the target population aged 18-69 were divided into two groups, based on categories of age (18-44 and 45-69) and gender (men and women).

In the subsequent lines the symbol "***l***" was used to denote the age group and "**m**" for gender.

Age group (*l*)	Gender (m)
$l = 1$ if age group (18-44)	m = 1 for men
$l = 2$ if age group (45-69)	m = 2 for women

For urban areas,

Define:

$$\delta_{ijknlm} = \begin{cases} 1 & \text{if } n^{th} \text{ selected respondent of the } k^{th} \text{ household of the } j^{th} \text{ CEB of the } i^{th} \\ & \text{ward belongs to age groups "} l \text{" and gender "m"} \\ 0 & \text{otherwise} \end{cases}$$

$\widehat{N}_{lm}$ = estimated number of persons of age group "l" and gender "m" if one person from the list of persons aged 18-69 was selected from household (l=1,2 and m=1,2)

$\widehat{N}_{lm}$ is obtained as

$$\widehat{N}_{lm} = \frac{1}{\hat{\theta}_{lm}} \sum_{overall\ possible\ values\ of\ i,j,k,n} HWT_{ijk} \times \delta_{ijknlm}$$

Where $\hat{\theta}_{lm}$ is the estimated group response rate.

Calibrated individual weight for urban areas

$$IWT_{ijklm} = \frac{N_{lm}}{\hat{N}_{lm}} \text{ x } HWT_{ijk}$$

Denoted by

N_{lm} = number of persons who belonged to gender "m" and age group "l" in the urban areas. (l = 1,2 and gender m =1,2)

Y_{ijkn} = the observed value of the study variable for the respondent "n" belonging to household "k", CEB j and ward i.

Estimate of the population total of gender group "m" and age group "l" is

$$\hat{Y}_{l,m} = \sum_{overall\ possible\ values\ of\ i,j,k,n} \delta_{ijknlm} \times IWT_{ijklm}$$

$$\hat{N}_{l.} = \hat{N}_{l1} + \hat{N}_{l2} \quad , \quad \hat{Y}_{l.} = \hat{Y}_{l1} + \hat{Y}_{l2} \ , \qquad l = 1, 2$$

$$\hat{N}_{.m} = \hat{N}_{1m} + \hat{N}_{2m} \ , \quad \hat{Y}_{.m} = \hat{Y}_{1m} + \hat{Y}_{2m} \ , \quad \text{m=1,2}$$

$$\hat{N} = \sum_{l=1}^{2} \sum_{m=1}^{2} \hat{N}_{l,m} \quad , \quad \hat{Y} = \sum_{l=1}^{2} \sum_{m=1}^{2} \hat{Y}_{l,m}$$

Estimate of the mean of the study variable for gender group m and age group l, $\hat{Y}_{lm}$ for and overall are

$\frac{\hat{Y}_{lm}}{\hat{N}_{lm}}$ $\frac{\hat{Y}_{m.}}{\hat{N}_{m.}}$ $\frac{\hat{Y}_{.l}}{\hat{N}_{.l}}$ $\frac{\hat{Y}}{\hat{N}}$ respectively

For rural areas,

Define:

$$\delta_{iknlm} = \begin{cases} 1 & \text{if n}^{th} \text{ selected respondent of the k}^{th} \text{ household of the i}^{th} \text{ village} \\ & \text{belongs to age group "}l\text{" and of gender "m".} \\ 0 & \text{otherwise} \end{cases}$$

N_{lm} = estimated number of persons of age group "l" and gender "m" if one person from the list of persons aged 18-69 was selected from household (l=1,2 and m=1,2)

$\hat{N}_{lm}$ is obtained as $\hat{N}_{lm} = \frac{1}{\hat{\Theta}_{lm}} \sum_{overall\ possible\ values\ of\ i,j,k,n} HWT_{ik} \times \delta_{ijknlm}$

Where $\hat{\theta}_{lm}$ is the estimated group response rates.

Calibrated individual weight for rural areas

$$IWT_{iklm} = \frac{N_{lm}}{\widehat{N}_{lm}} \times HWT_{ik}$$

Denoted by

N_{lm} = number of person of gender m belonging to age group "l" in the rural areas of the population

(l = 1,2 and gender m =1,2)

Y_{ikn} = the observed value of the study variable for the respondent "n" belonging to household "k" of village i.

Estimate of the population total of gender group m and age group "l"is

$$\widehat{Y}_{lm} = \sum_{overall\ possible\ values\ of\ i,j,k,n} \delta_{ijknlm} \times y_{ikn} \times IWT_{ijklm}$$

Estimate of the mean of the study variable for age-gender group "l'' and "m", gender group "m", age group "l'' and overall mean can be obtained.

Individual sample weights for adolescents aged 15-17 years followed the similar procedure as described above.