Weighting for the India Human Development Survey (IHDS) Wave 3 data

Sharan Sharma, Steven Heeringa, and Sonalde Desai

1. Introduction

Almost all analyses using IHDS data will require the analyst to use weights provided in the IHDS public use dataset. This is necessary to prevent biases in inferences stemming from unequal selection probabilities and differential non-response. The objective of this document is to provide an understanding of how IHDS-3 weights have been developed. This document does not address how the weights are to be used during analysis; West et al. (2025) should be consulted for this purpose.

We first describe the IHDS sample design and survey instruments. We then discuss the development of IHDS-3 household weights, beginning with weights for cross-sectional analysis and then for panel analysis. After this, we describe the development of cross-sectional and panel weights for individual-level analysis. Two special cases of individual weights are then described - that of Eligible Women and then Migrants.

2. The IHDS design

The IHDS survey population comprises all private households and individuals in rural and urban India except the islands in Andaman and Nicobar, Lakshadweep. The first IHDS wave was conducted in 2004-2005 by interviewing 41,554 households. The sampling frame was based on an earlier enumeration and sampling exercise conducted by the National Council of Applied Economic Research (NCAER), which was suitably expanded. States and urban/rural strata were the two principal axes of stratification. Within rural strata, districts were grouped into three approximately equal-population strata based on an index of development. The rural sample comprised 26,734 households (64%) from 1504 villages. Villages were sampled within each stratum using simple random sampling and then households chosen within each village with a sample size proportional to the household population in that village. The urban sample consisted of 14,820 (36%) households drawn from 970 urban areas (towns and cities). The urban areas were chosen PPS (Probability Proportionate to (household population) Size) and then an equal number of households chosen within each area. Given the above sampling design, both rural and urban

strata resulted in an approximately equal-probability sample within each state. Weights for IHDS-1 were simply the inverse of the sample units' selection probabilities.

The second wave of IHDS was conducted between 2001-2012 and collected data from 34,643 IHDS-1 households (representing an 83.3% AAPOR RR1 re-interview response rate). To partially compensate for attrition vis-a-vis IHDS-1, 2,134 households were recruited afresh. Further, 5,375 split households were interviewed - these were households that were formed due to member/s moving out of the root household into new housing units in the same village or urban locality. In all, 42,152 households were interviewed in IHDS-2.

The third (present) IHDS wave was conducted between March 2022 and June 2024 and attempted to follow all households who responded to at least one of IHDS-1 and IHDS-2. A total of 47,842 households reported data in IHDS-3 of which a substantial 82% (39,223 households) have their origins in IHDS-1, a good proportion considering a time span of close to two decades between the first and third waves.

To compensate for attrition of IHDS-2 households, 7,377 new households (15.4% of the IHDS-3 sample) were recruited. Operationally, we wanted to avoid a situation where field staff may not put in their full effort in obtaining cooperation from panel households and just choose to just recruit substitute households. Therefore, as a blanket rule based on early estimates of non-response, we asked staff to always recruit 2 new households in rural PSUs and 5 new households in urban PSUs. Given operational constraints, there was no sampling frame constructed separately for this purpose but supervisors would give a random starting point within each PSU from where field staff would start attempting to recruit the new households. Given this procedure, calculation of selection probabilities was not possible for these new households but fieldwork was monitored closely and discussed in weekly field staff meetings. The distribution of the IHDS-3 household sample by state-urban/rural strata is given below in Table 1. Some of the reporting units in Table 1 are not technically states (e.g. Daman & Diu) but we refer to these as "states" for simplicity.

Table 1: IHDS-3 sample sizes by states/reporting geographies and urban status. The geographies are arranged in ascending order of IHDS codes.

State	Rural	Urban
1. Jammu & Kashmir	552	456
2. Himachal Pradesh	1,302	341
3. Punjab	1,277	461
4. Chandigarh	0	52
5. Uttarakhand	330	208
6. Haryana	1,780	295
7. Delhi	28	930
8. Rajasthan	2,186	920
9. Uttar Pradesh	3,310	1,361
10. Bihar	1,303	583
11. Sikkim	22	81
12. Arunachal Pradesh	107	72
13. Nagaland	146	90
14. Manipur	92	70
15. Mizoram	92	48
16. Tripura	231	74
17. Meghalaya	115	56
18. Assam	1,012	383
19. West Bengal	1,816	1,341
20. Jharkhand	643	446
21. Orissa	1,752	549
22. Chhattisgarh	1,233	358
23. Madhya Pradesh	2,981	754
24. Gujarat	1,109	707
25. Daman & Diu	56	0
26. Dadra & Nagar Haveli	58	18
27. Maharashtra	2,471	1,129
28. Andhra Pradesh	1,433	735
29. Karnataka	3,078	1,179
30. Goa	91	49
32. Kerala	632	844
33. Tamil Nadu	892	1,037
34. Pondicherry	54	31
Total	32,184	15,658

3. IHDS-3 fieldwork and instruments

A PSU was typically covered by five field staff consisting of two male-female interviewer pairs and one supervisor. The interviewer pairs typically administered instruments at the household while the supervisor administered instruments at the community level that required visits to local facilities such as schools.

After obtaining cooperation from the household, the male interviewer (usually) began administering the Roster instrument mainly dealing with household composition to the head of the household or another responsible household member. After the Roster was administered, the male interviewer continued to interview the Roster respondent by administering an "Income and Social Capital" (or "household") instrument, a core instrument which collects data spanning diverse topics such as internal migration, farming, jobs, income, etc. While many questions in this instrument are at a household level, e.g., ownership of farm land, some questions are also at an individual level, e.g., government benefits received by each household member.

While the male interviewer was administering the household instrument, in parallel the female interviewer would typically administer an 'Education and Health' (EDH) instrument to an 'Eligible woman' (EW) of the household covering questions on education of the household members, water and sanitation, morbidity, etc. Similar to the household instrument, the EDH instrument contains questions at the household and individual level.

After the EDH instrument, based on automatic in-built rules in the Roster instrument (described later in Section 9) the female interviewer administered an EW instrument to specific women in the household. The EW instrument has questions on the EW's health beliefs, marital history, fertility, etc.

Children between the ages of 8 and 11 years in the household are administered reading and mathematics tests by either the male or female interviewer. If children were in school when the interviewers visited the household, they made an appointment and visited the household again. Finally, anthropometric measurements (heights and weights) were sought to be taken for all members of the household but with a special focus on EW, children less than 18 years of age, and the primary respondent to the household questionnaire. Just as with the learning tests, not all

members might have been available for these measurements and interviewers made multiple attempts while in the PSU to obtain these measurements.

In all, except for the EW instrument which may be administered to *sampled* household members, all other individual-level instruments were designed to be for all members of the household in the target population.

4. IHDS-3 Household Cross-sectional Weights

IHDS-3 cross-sectional (or wave-specific) weights are used when only data from IHDS-3 are analyzed. Since the IHDS-3 sample consists of households from previous waves, we use this as a starting point for our weight construction. Figure 1 depicts four components of the IHDS-3 sample with reference to the previous IHDS waves.

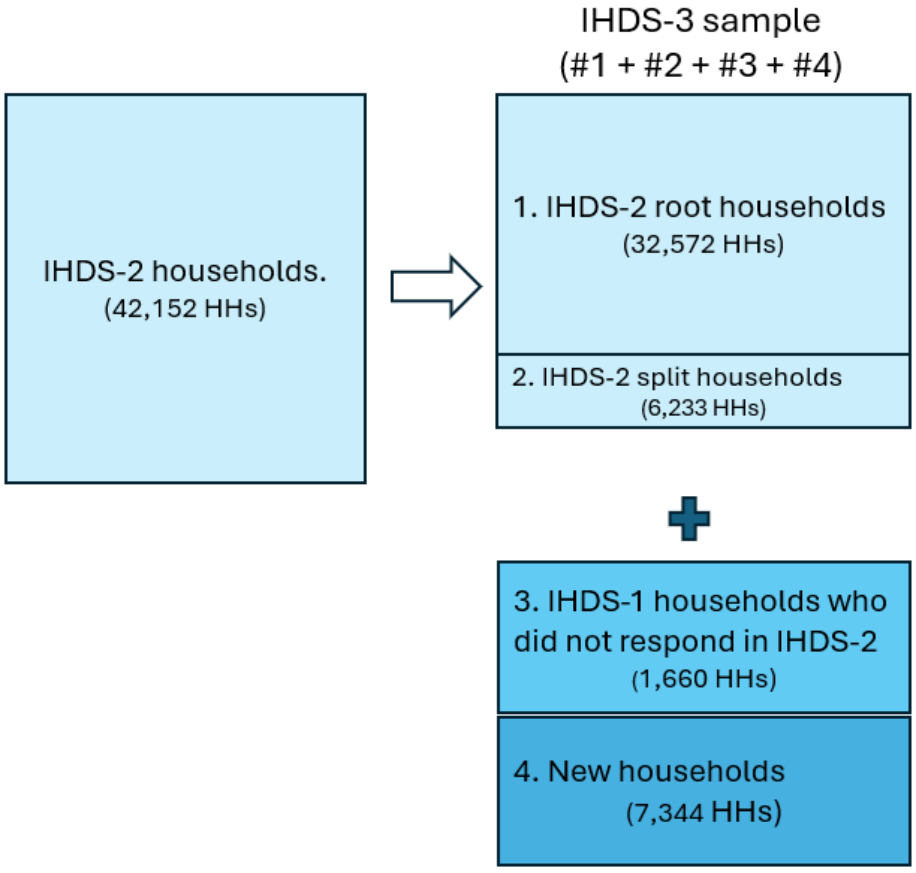

Figure 1. IHDS-3 sample components.

Figure 1 shows that a large (78%) component (component #1) of the IHDS-2 households also responded in IHDS-3 forming 68% of the IHDS-3 sample. Also, some of these IHDS-2 households had split into new household units by IHDS-3 (component #2) yielding a further 6,233 households (13% of the IHDS-3 sample). Component #3 comprises 1,660 households who previously reported in IHDS-1 but not in IHDS-2, comprising a relatively small 3.5% of the IHDS-3 sample . Finally, 7,377 households (component #4) were recruited afresh to compensate for panel attrition and to infuse fresh households into the sample. This forms 15.4% of the IHDS-3 sample. Overall, two-thirds of the 2470 IHDS PSUs have a larger sample size in IHDS-3 than in IHDS-2 (Figure 2).

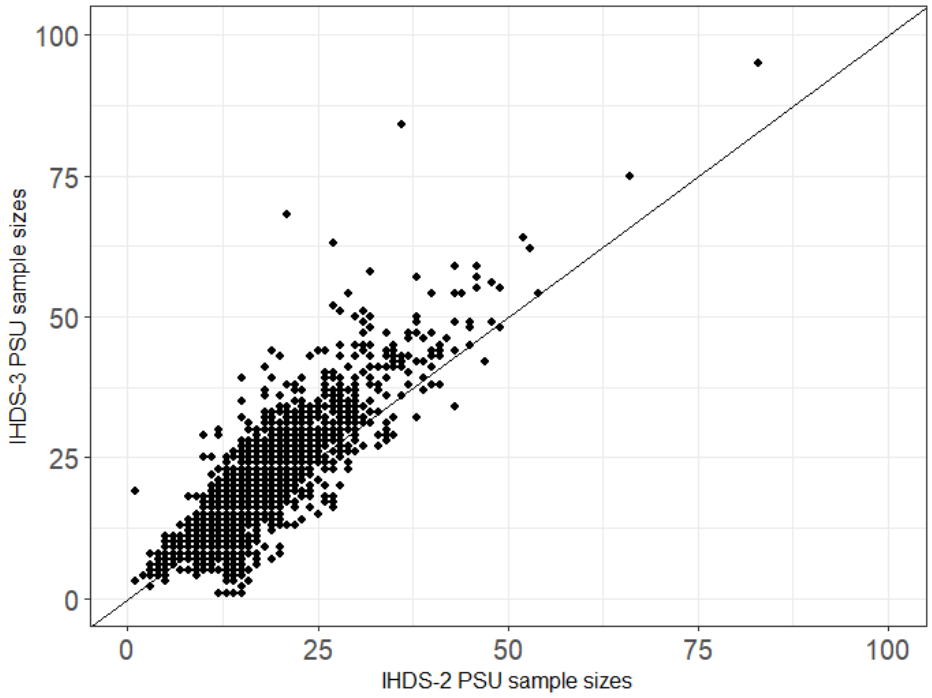

Figure 2. IHDS-2 sample sizes (horizontal axis) versus IHDS-3 sample sizes (vertical axis) for each PSU in the IHDS-3 sample (dots). The diagonal line is the 45 degree line at which the sample sizes for the two waves are equal for a PSU.

The construction of the cross-sectional household weights proceeds as follows:

A. Step 1: Base weights

Given IHDS sampling procedures, sample households in the four components in Figure 1 are assumed to have the same selection probabilities within a PSU. The cross-sectional base weight for household □ in PSU □ is therefore calculated as:

$$\square_{\square\square}{}^{\square\square\square\square\_\square\square\square\square 3} = \frac{\Sigma_{\square'} \ \square_{\square'\square}{}^{\square\square\square\square 2}}{\square_{\square}{}^{\square\square\square\square 3}}$$

i.e., the sum of weights of all □′ households in PSU □ who reported data in IHDS-2 divided by the sample size for PSU □ in IHDS-3 (sum of the four sample component sizes in Figure 1). Eight PSUs were not worked in IHDS-2 but in IHDS-1 and IHDS-3. We took the IHDS-1 design weights in those cases. A summary comparison of IHDS-2 and IHDS-3 weights is given in Table 2 below.

Table 2. Comparison of IHDS-2 and IHDS-3 weight summary measures

	IHDS-2	IHDS-3
Minimum	154.1	154.1
1st quartile	3,480	2,909
Median	4,953	4,565
Mean	6,898	6,502
3rd quartile	8,110	7,694
Maximum	152,767	135,792
Number of PSUs with a weight 3 times greater than the median	117	122

The increase in sample sizes in IHDS-3 has meant a general decrease in the magnitude of weights. However, 122 PSUs have large household weights (defined here as being more than three times the median weight).

B. Step 2: Weight trimming

Large weights can increase variance in statistical estimates and therefore need to be trimmed. To do this we take an approach advocated in Van de Kerckhove et al. (2014) that links the maximum weight to the increase in variance due to weighting (Kish 1992):

$\square_{\square\square\square\square\square\square} = 3.5 * \sqrt{1 + \square\square^2(\square_{\square})} * \square\square\square\square\square\square(\square_{\square})$, where CV is the "coefficient of variation".

Given that the number of PSUs and household sample sizes vary by strata, weight trimming was undertaken for each state x urban/rural stratum separately. Small states/reporting geographies were grouped for this purpose (see Appendix A). Of the total 42 state groups x rural/urban strata, 22 required trimming. We compute the implicit percentile at which the weights were trimmed; values range from the 94th percentile to the 99th percentile with 16 of the 22 state groups having their weights implicitly trimmed at the 98th percentile or above. Of the 47,842 households, only 536 households required weight trimming.

We examined the sum of weights in each stratum after trimming. While most of the 22 strata groups lose less than 2% of the sum of weights due to trimming, five strata groups lost more than 6% of the sum of weights. Since this can impact interstate comparability during analysis, the difference between the sum of the original base weights and the sum of the trimmed weights was distributed back across all households equally in that stratum.

C. Step 3: Weighted sample checks

We compared the profiles of the weighted IHDS-2 sample and the IHDS-3 sample for three key variables: farm ownership, religion, and caste. These variables were chosen since they are frequently used in analyses, are correlated with many key outcome variables of interest, and are stable across time at a state x urban/rural level.

Figures 3 - 6 plot IHDS-2 and IHDS-3 weighted estimates for categories of the three analysis variables. Each point in the plot is a state x urban/rural stratum. The horizontal and vertical axes are the category proportions in IHDS-2 and IHDS-3 respectively. The diagonal band represents a 10 percentage point difference between the wave proportions which may be thought of as an acceptable range of difference. The figures show that weighted estimates of the proportion of farm owners (Figure 3), Muslim household heads (Figure 4), "general category caste" household heads (Figure 5), and scheduled caste/tribe caste household heads (Figure 6) are quite consistent between IHDS-2 and IHDS-3. Even the few points that are outside the diagonal band are close to

the band's outer limits. Delhi-rural shows up outside the diagonal band in Figures 3,4, and 5 but it has a very small sample size as shown in Table 1 (just 28 households).

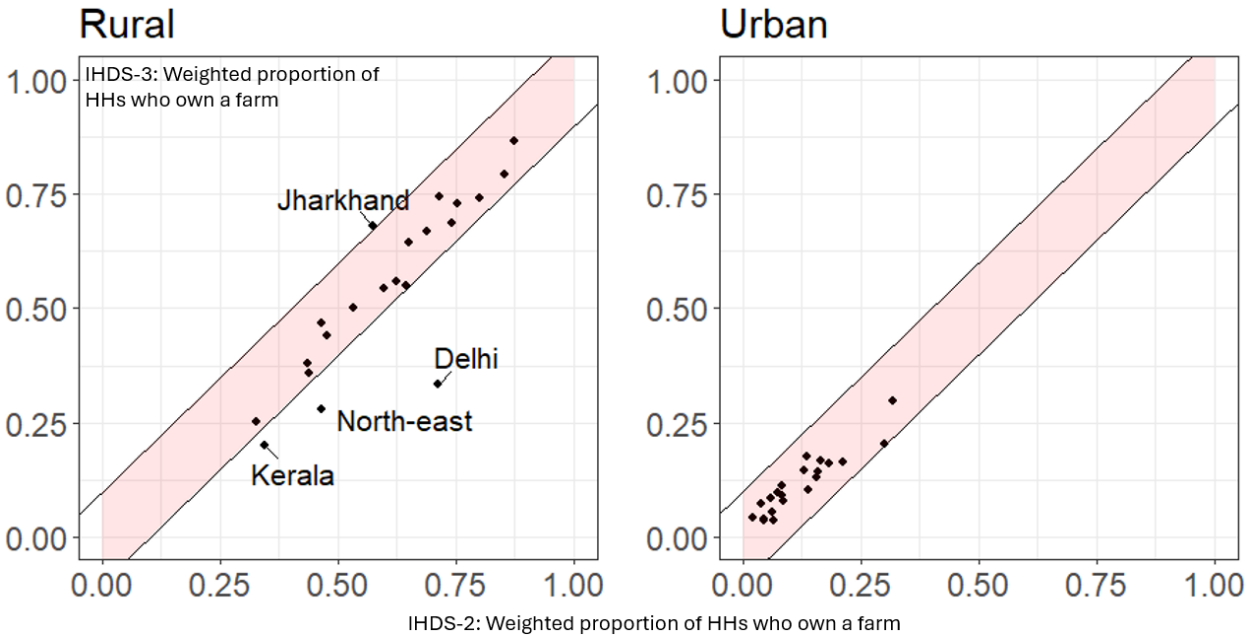

Figure 3: Comparison of IHDS-2 and IHDS-3 on weighted estimates of farm ownership

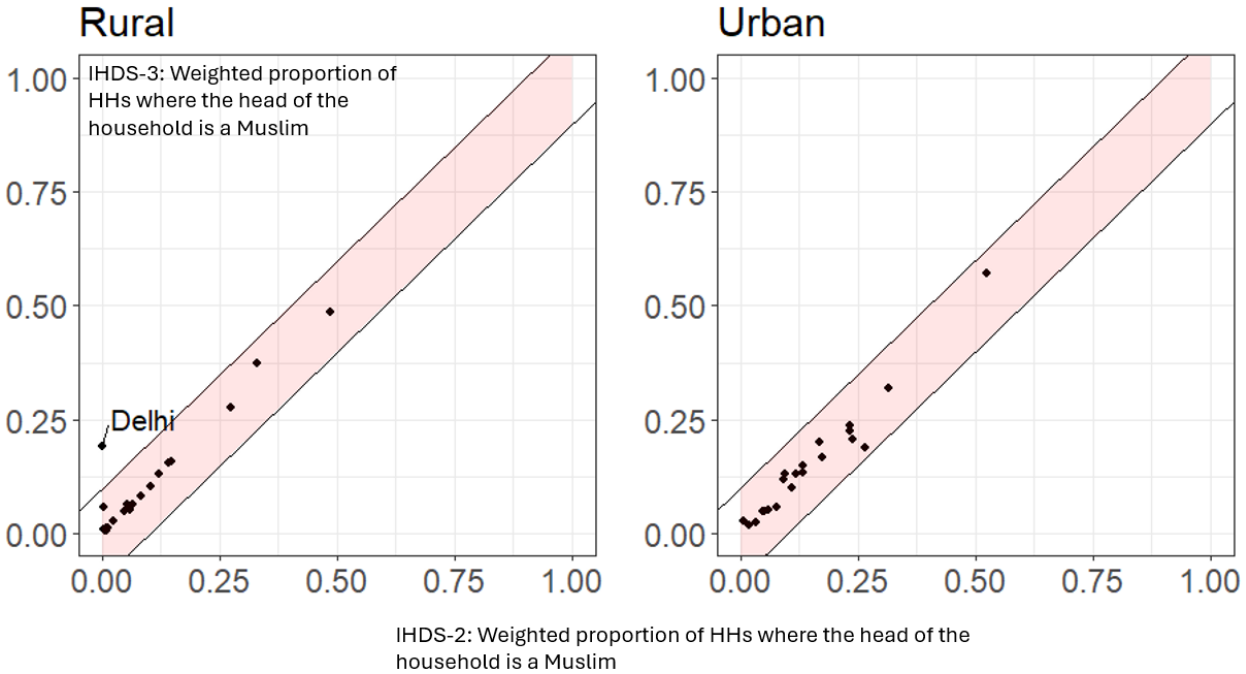

Figure 4: Comparison of IHDS-2 and IHDS-3 on weighted estimates of the proportion of households where the head of the household is a Muslim

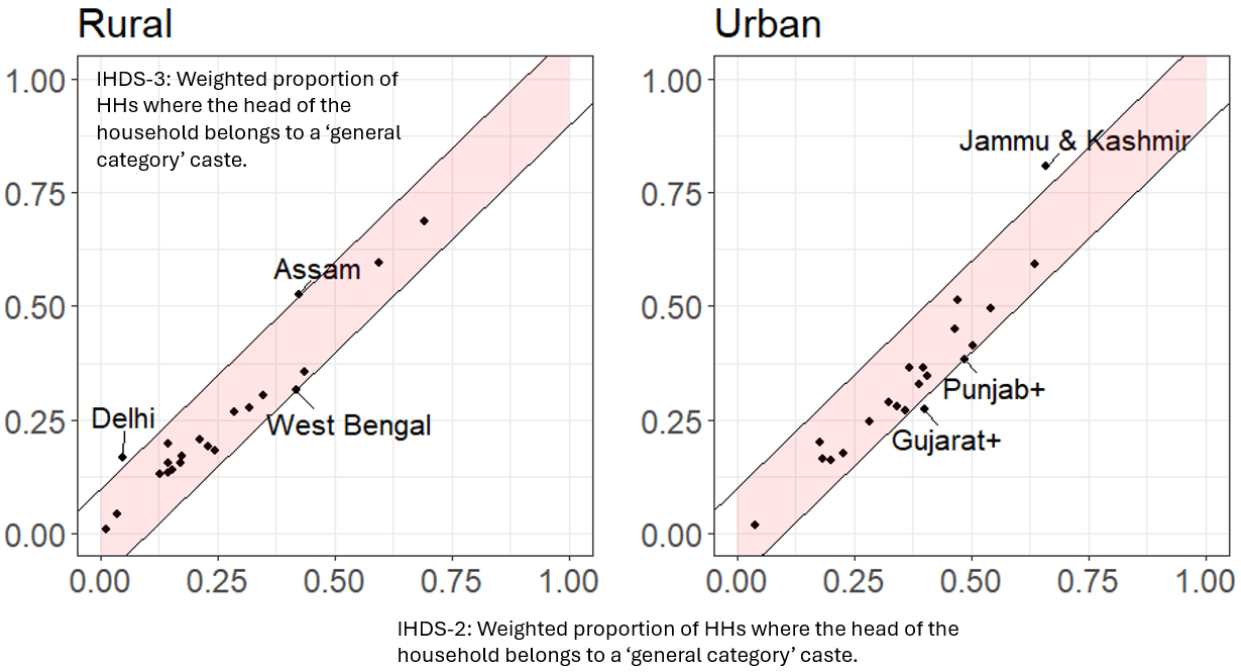

Figure 5: Comparison of IHDS-2 and IHDS-3 on weighted estimates of the proportion of households where the head of the household belongs to a 'general category' caste

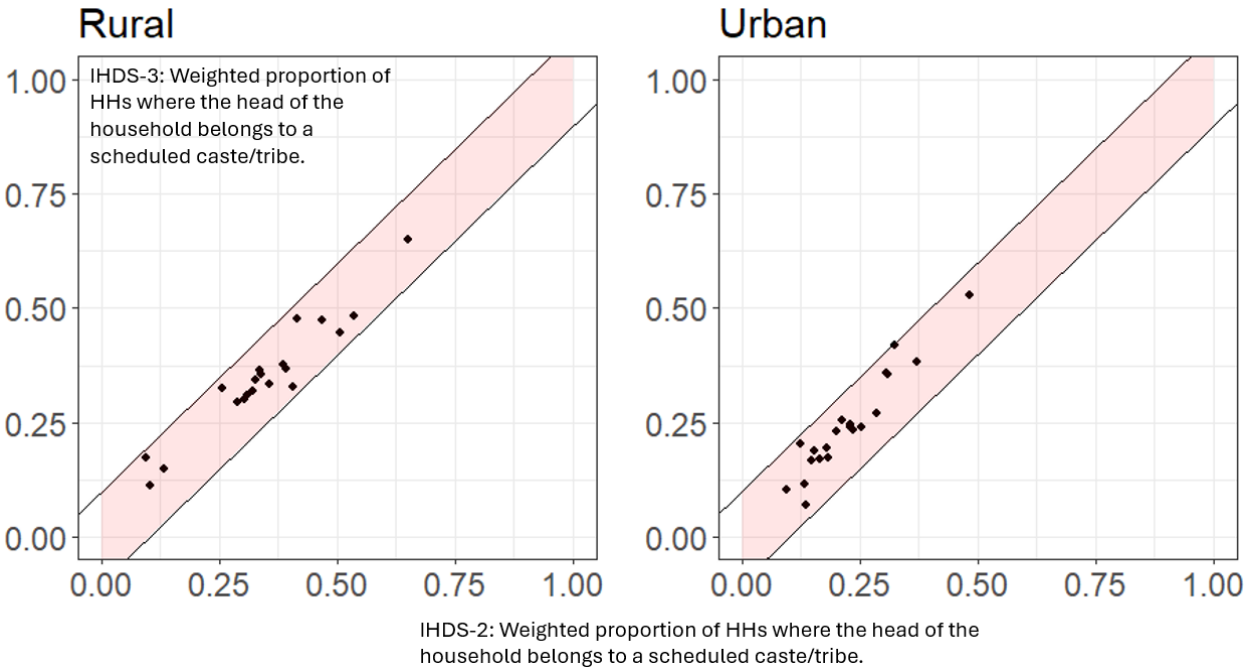

Figure 6: Comparison of IHDS-2 and IHDS-3 on weighted estimates of the proportion of households where the head of the household belongs to a scheduled caste/tribe

Finally, Figure 7 compares the distribution of IHDS-2 and IHDS-3 cross-sectional weights and shows that IHDS-3 weights are generally smaller on average (due to the increased sample sizes) and smoother. The smoother weights will translate into smaller standard errors of estimates.

Figure 7: Density plot of IHDS-2 and IHDS-3 cross-sectional weights. Horizontal axis is truncated at 20,000.

5. IHDS-3 Household Panel Weights

A. Panel household base weights

Depending on the analysis, we have two types of base panel weights:

- For analyses involving waves 2 and 3 (IHDS-2-3), the analytic sample is made of only components #1 and #2 of Figure 1. The design weights for these households are the same as IHDS-2 design weights. Households that split-off (component #2 of Figure 1) in IHDS-3 get the same design weight as the root household (component #1 of Figure 1). Note that the designation of a household as 'root' and 'split' is arbitrary.
- For analyses involving all three IHDS waves (IHDS-1-2-3), the design weights will be the IHDS-1 design weights and as earlier, split households get the same weight as the root household.

We do not develop panel weights for analysis involving only IHDS-1 and IHDS-3; as Figure 1 shows (component #3), very few households report data in IHDS-1 and IHDS-3 but not in IHDS-2. We describe below the development of weights for the IHDS-2-3 panel analysis. The IHDS-1-2-3 panel weights proceed in exactly the same fashion and are therefore not described in this document.

B. Adjustments for non-response

Ideally we would have the entire IHDS-2 household sample reporting data in IHDS-3. However, some households did not respond in IHDS-3, either because they were non-contactable (e.g., away on vacation) or did not want to take part in the survey. Response rates (AAPOR RR1 response rates) for the various states and urban status' are displayed in Figure 8 below.

The average response rate across reporting geographies for the rural stratum was 84% with 28 of the 33 states/reporting geographies registering a response rate greater than 75%. The lowest rural response rate was in Kerala at 66%. Urban response rates were lower compared to the rural stratum; the average urban response rate was 66%; 32 of the 33 states had a response rate greater than 60%. The minimum urban response rate is in the city of Chandigarh at 39%.

Non-response not only results in a lower sample size but if the non-responding homes have a different mean value for an analysis variable than the responding households, our inferences would be biased since the unadjusted estimate is computed only on the responding households. Given that we do not have substantive data for the non-responding households, we make use of auxiliary variables which are household characteristics available for both responding and non-responding households. This helps us construct non-response weights to adjust for possible non-response bias. Auxiliary variables in our case are in the form of several demographic variables captured for the IHDS-2 households.

For both reduced bias and variance of a statistic, the auxiliary variables should be correlated with the response propensities (□) of the households and the substantive analysis variable (□) . If the auxiliary variable is only related to □ and not to □, then the weights will not reduce bias but will actually result in increased variance of the survey estimate (Little and Vartivarian 2005). Since IHDS reports hundreds of variables, it is difficult to arrive at a set of auxiliary variables that will

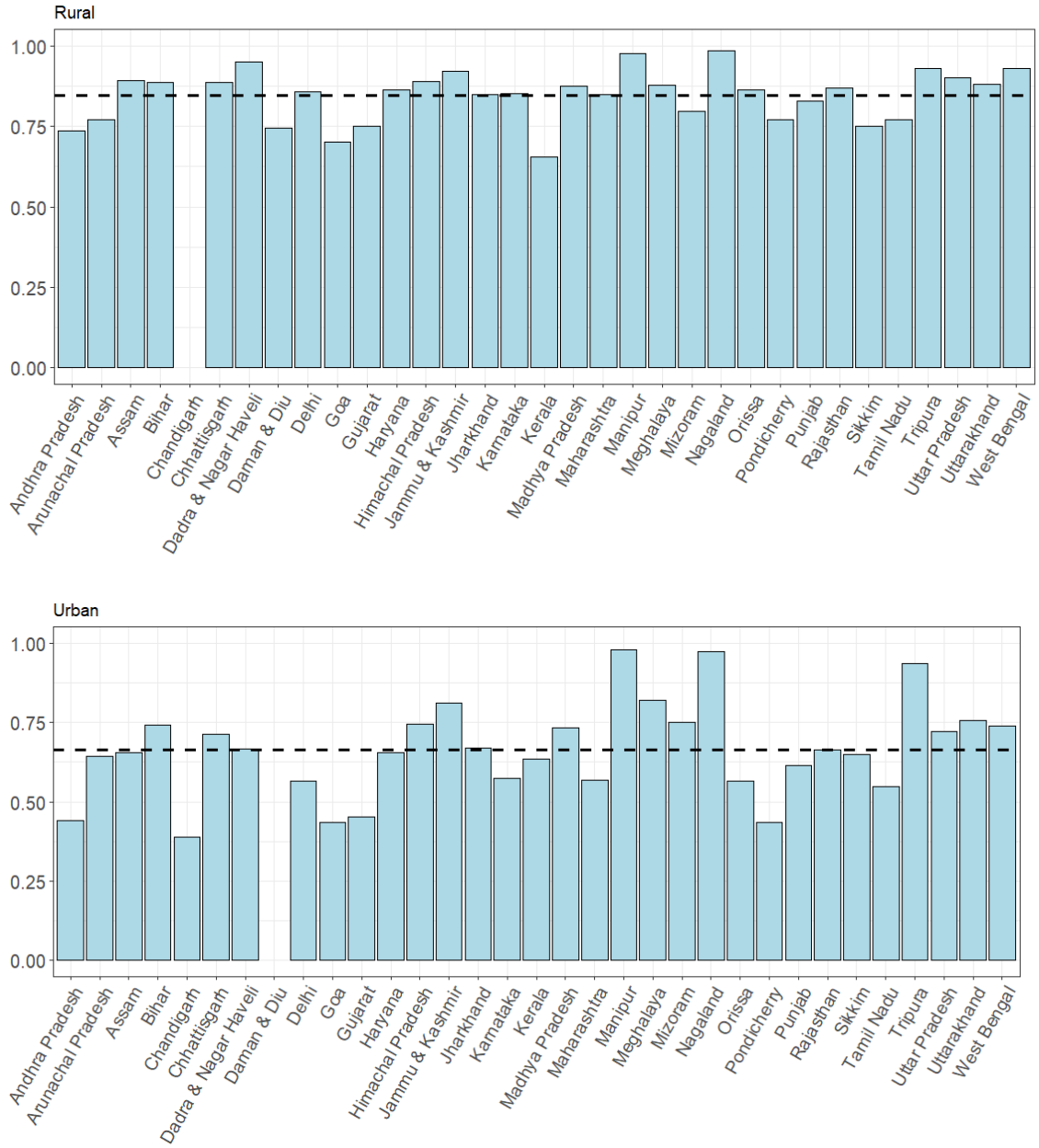

Figure 8: IHDS-3 response rates for different states on the IHDS-2 sample base. The top panel shows response rates for the rural stratum while the bottom panel is that of the urban stratum. The dashed horizontal line is the average response rate across states.

correlate with all analysis variables. Therefore, based on our knowledge of frequent IHDS analyses, we decided to use seven variables for non-response adjustment. The sample distribution of the variables and the unweighted response rate for the full IHDS-3 sample within each category of the adjustment variables is shown in Table 3 below. We see that response rates for some variables across their categories are quite different. For example, the response rate for households

with the highest asset count is 64% compared to a 79% response rate for households with the lowest asset count.

Table 3: Sample distribution of variables used for non-response adjustment and response rates within variable categories.

Religion-caste	IHDS-2 Sample %	IHDS-3 Response rate %
Hindu - General category	22%	72%
Hindu - OBC	34%	78%
Muslim	11%	76%
Other*	33%	81%

**includes scheduled caste/tribes and other religions*

Household size	IHDS-2 Sample %	IHDS-3 Response rate %
1-2 persons	13%	59%
3-4 persons	36%	75%
5+ persons	51%	83%

HH has a child ?	IHDS-2 Sample %	IHDS-3 Response rate %
No	37%	70%
Yes	63%	82%

HH owns a farm ?	IHDS-2 Sample %	IHDS-3 Response rate %
No	55%	70%
Yes	45%	86%

# assets owned	IHDS-2 Sample %	IHDS-3 Response rate %
0 - 5	7%	79%
6 -15	41%	85%
16 -20	25%	78%
21+	26%	64%

Toilet type	IHDS-2 Sample %	IHDS-3 Response rate %
None	45%	85%
Pit or Septic	45%	73%
Flush	10%	63%

HH head is a graduate ?	IHDS-2 Sample %	IHDS-3 Response rate %
No	80%	80%
Yes	20%	66%

Since non-response levels are different for the 42 state-urban/rural strata with potentially different underlying mechanisms, the non-response adjustments were done separately for these strata as follows:

1. An indicator household response variable was attached to the IHDS-2 dataset, taking a value of 1 if the IHDS-2 household responded in IHDS-3 and 0 if not.

2. A logistic model is fit to the binary household response indicator using the above variables as main effects. The farm ownership and toilet type variables were used only for the rural models. After examining the initial fit, variables were removed from the model if they were not (statistically) significant predictors, evaluated at $\alpha = 0.05$. This results in a different model functional for each stratum. Delhi rural was combined with Delhi urban for sample size reasons.
3. The predicted response propensities ($\hat{\square}$) were sorted and visually examined. For an overwhelming majority of strata, the $\hat{\square}$ were well separated leading to the forming of natural non-response adjustment classes. For example, the left panel of Figure 9 shows the sorted predicted probabilities for urban Uttar Pradesh (n=1301) suggesting the formation of four adjustment classes: < 0.5, 0.5-0.6, 0.61-0.8, > 0.8. However, we set a minimum sample size of 50 households for each class and the < 0.5 class had only 34 cases. This class was therefore combined with the 0.5-0.6 class. In a few cases, the predicted response propensities were not as clearly separated such as for urban Punjab (n=1177) seen in the right panel of Figure 9. In such cases, we looked for discontinuities in the distribution to form the classes. In this case, the four classes formed were < 0.7, 0.7-0.8, 0.81-0.84, > 0.85.

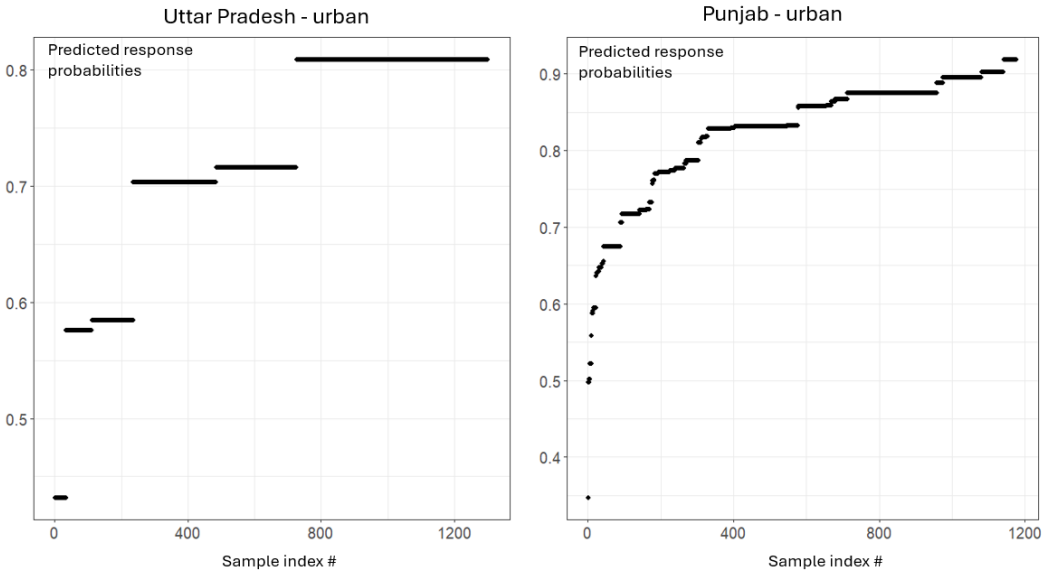

Figure 9. Predicted response propensities for urban Uttar Pradesh (left) and urban Punjab (right)

4. Within each class, the non-response adjustment for adjustment class □ was computed as: $\phi_{\square} = \sum \hat{p}_{i_c}/n_c$, i.e., the average estimated propensity (Valliant et al 2013, p.330).

C. Final panel weights

For each IHDS-2 household □ that responded in IHDS-3 having design weight □□ and in non-response adjustment cell □, the final panel weight was computed as: □□ = □□ ∗ (1/ □□). The weights were examined and trimmed following the same procedures used for the cross-sectional weights described earlier. All IHDS-3 split households get the same panel weights of the 'root' IHDS-2 household. In all, 38,805 households (components #1 and 2 in Figure 1) will have associated panel weights.

6. IHDS-3 Individual Cross-sectional weights

Given that there is no within-household sampling per se, for all analyses (except those of the eligible woman instrument) all individuals in the home inherit the same weight as the household weight.

7. IHDS-3 Individual Panel weights

8. Eligible women (EW) Cross-sectional weights

The target population of Eligible Women (EW) is defined as ever-married female household members between 15 and 49 years of age (both ages included). The term "ever-married" means that women who are now widowed or divorced but were married at some point are also part of the target population. EW in IHDS sample households are interviewed by female interviewers by administering an instrument containing questions on gender relations, marital history, birth history, etc.

The cross-sectional weight for EW respondent □ for IHDS responding household □ is:

$$□_{□□}^{□□\_□□} = □_{□}^{□□\_hh} ∗ □_{□} ∗ (1/□_{□□}) \quad \text{----- (2)}$$

where $□_{□}^{□□\_hh}$ is the cross-sectional household weight, □□ is the probability of selecting the EW among all EW in the household, and $1\backslash\phi_{ij}$ is a non-response adjustment. The latter two components are explained below.

a. EW selection probabilities

If all EW were sampled in an interviewed household, then there would be no individual selection probabilities to consider. However:

- To enhance the longitudinal utility of the data, EW from previous waves (panel EW) were selected for the IHDS-3 EW interview even if they were more than 49 years old. This consideration applied, of course, only to panel households and not households recruited afresh in IHDS-3.
- Operational considerations meant that we would not be able to interview more than three EW in a household.

To meet these considerations, once the Roster instrument was completed, EW sample lines were 'spawned' for *all* panel EW. Then, if new EW members were available and the maximum limit of three EW was not yet reached, new EW were sampled from the available pool using simple random sampling without replacement and interview lines spawned for such samples.

Table 4 shows a crosstab of the number of old and new EW in an IHDS-3 interviewed household. We find that (a) The number of panel EW never exceeds 2, and (b) Only 278 households needed to have a sampling of an EW member (shaded cells in Table 4) - this represents less than 2% of EW sample members in the target population.

Given the small proportion of households where some sampling of EW needed to be done, we ignore the within-household selection probabilities in such cases. For most analysis, this is expected to lead to only a small bias but might result in lesser mean squared error due to reduced variance, especially for subpopulation analysis. Thus, we set $\square_{\square} = 1$ in Equation 2 for all EW.

Table 4: Cross-tab of the number of IHDS-3 households with the number of panel EW (rows) and new EW (columns). The shaded cells are those where new EW members were sampled.

	#new EW (not interviewed in IHDS2) →						
#panel EW ↓	0	1	2	3	4	5	6
0	-	12,886	1,121	129	14	2	1
1	19,559	6330	1,142	137	16	2	1
2	1,131	389	85	16	3	1	0

b. Non-response

Not all EW selected were able to be interviewed. The response rates for different geographies are plotted in Figure 10 and Figure 11 for the rural and urban strata respectively. The response rates were quite high (dashed line in Figures 10 and 11) - at 90.2% for the rural stratum and 88.4% for the urban stratum and barring a few exceptions, we do not see a lot of between-state variation in response rates.

These response rate calculations - and in general, the development of EW weights - are done only for those members who meet the target population definition. That is, members who were administered the EW instrument but were more than 49 years old on the interview date do not have an EW-specific cross-sectional weight associated with them. Such individuals would be analyzed using the usual individual cross-sectional weights (Section 6 above). To understand this more, Table 5 shows the sample and respondent counts of EW members based on whether these members were panel EW or new recruitments and if the panel members continue to fit the target population definition. The 37,144 respondents in the light yellow cells are those which will have an associated EW weight. The 11,073 panel EW members who no longer fit the EW target population definition will have a zero EW weight but the data file will contain the regular individual cross-sectional weight for more general analysis.

Table 5. IHDS-3 EW sample and respondent composition by panel and new recruitments and whether the sample is in the target population definition.

	New EW	Panel EW	
		Within Target Population definition (<50 years)	Outside Target Population definition (>49 years)
1. EW sample	19,918	21,491	12,662
2. # respondents	17,365	19,779	11,073
3. Response rate	87%	92%	87%

Table 6 displays the response rates by different household and individual variables for the full sample. These variables were chosen for being potentially associated with both, response propensities and common analysis variables. For example, 'number of children' was selected since EW with more children may be busier and therefore have lower propensities. Moreover, two EW with different numbers of children would likely differ on behaviors such as contraceptive usage.

At a full sample level, we do not see much variation in response rates by caste-religion, status as a housewife, and access to a mobile phone but these do show differences for specific states (not shown in this document). In general, lower response rates are associated with richer households (more assets) and households with multiple EW. Eligible women who are younger, have fewer children, and are less educated are associated with lower response rates. Given this evidence, we account for differential non-response by modeling the binary response indicator by the variables in Table 5. This modeling is conducted separately for each state group. Given that response rates are quite high in general and most states do not exhibit a large urban-rural disparity we do not model each urban/rural stratum separately within each state but just use urban/rural as a main effect in our models.

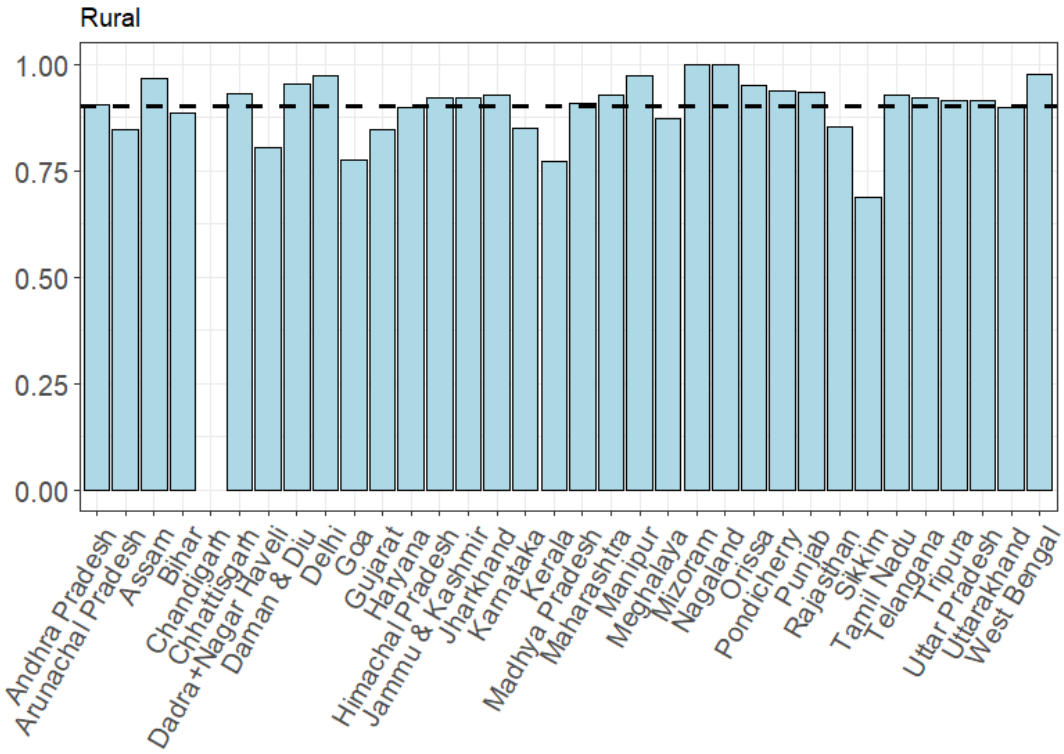

Figure 10: EW response rates by state for the rural stratum. The dashed horizontal line is the mean unweighted response rate across states.

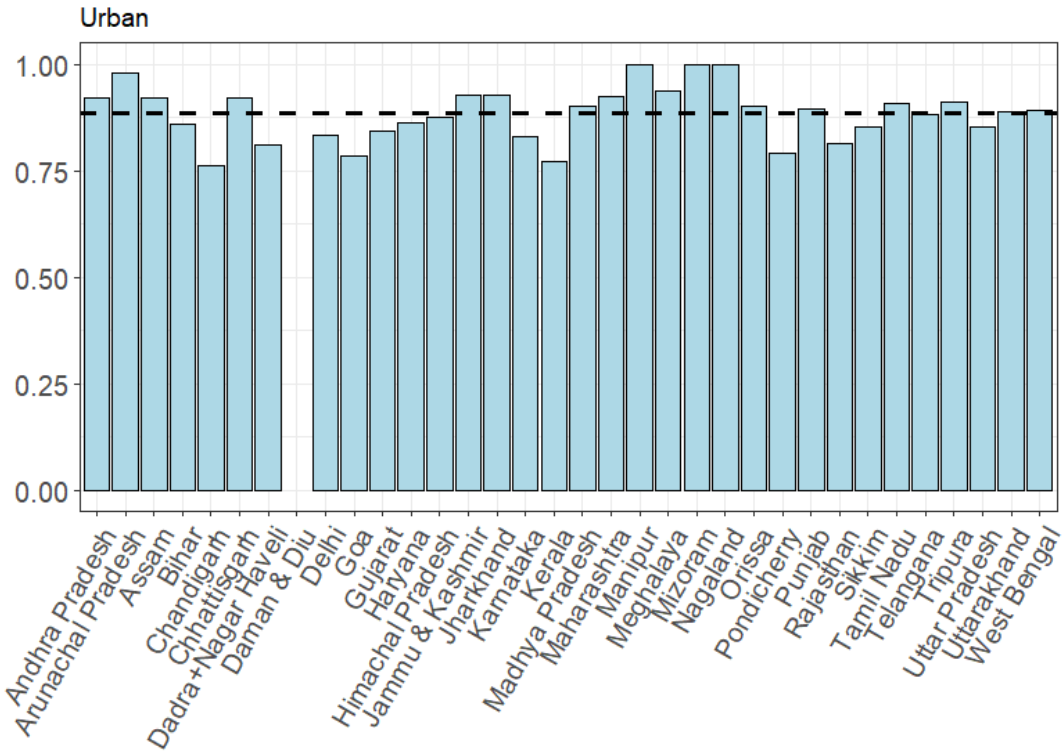

Figure 11: EW response rates by state for the urban stratum. The dashed horizontal line is the mean unweighted response rate across states.

Table 6: Eligible Women (EW) sample distribution of variables used for non-response adjustment and response rates within variable categories.

Religion-caste	Sample %	Response rate %
Hindu - General category	18%	89%
Hindu - OBC	34%	90%
Muslim	14%	88%
Other*	34%	90%

**includes scheduled caste/tribes and other religions*

# assets owned	Sample %	Response rate %
0 - 4	4%	91%
5 -10	23%	92%
11 - 14	37%	90%
15-19	27%	89%
20-34	9%	86%

#EW lines spawned	Sample %	Response rate %
1	74%	92%
2	22%	83%
3	4%	80%

# Children	Sample %	Response rate %
0	14%	84%
1	27%	88%
2	37%	91%
3	15%	92%
4 and above	7%	94%

Age	Sample %	Response rate %
15 -24	13%	84%
25 - 29	17%	86%
30 - 34	18%	89%
35 - 39	19%	92%

Education	Sample %	Response rate %
No schooling	25%	82%
Primary schooling	14%	91%
Middle schooling	17%	88%
Secondary schooling	19%	92%

40 - 49	33%	93%

Higher secondary schooling	12%	91%
College and above	13%	92%

Is a housewife ?	Sample %	Response rate %
Yes	68%	89%
No	32%	90%

#Access to a mobile	Sample %	Response rate %
Yes	70%	90%
No	30%	89%

The non-response adjustments are computed following the procedure laid out in Section 5.B. After this, weights were computed using Equation 2 and the procedure explained in Section 5.C. These weights were trimmed as described in Section 4.B and constitute the final weights.

References

West, B.T. , Heeringa, S. G., Berglund, P. A.(2025). *Applied Survey Data Analysis. 3rd Edition*. Chapman and Hall/CRC.

Kish, L. (1992). Weighting for unequal Pi. *Journal of Official Statistics*, *8*(2), 183.

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means?. *Survey Methodology*, *31*(2), 161.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples* (Vol. 1). New York: Springer.

Van de Kerckhove, W., Mohadjer, L., & Krenzke, T. (2014). A weight trimming approach to achieve a comparable increase to bias across countries in the for the international assessment of adult competencies. *Joint Statistical Meetings*, 655-666. Accessed at: http://www.asasrms.org/Proceedings/y2014/files/311170_87007.pdf

Appendix A. State/reporting geography groupings.

Original state/reporting geography	Grouped into…
Chandigarh	Punjab
Uttarakhand	Uttar Pradesh
Sikkim, Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura, and Meghalaya	North-east (new grouping created)
Daman & Diu and Dadra & Nagar Haveli	Gujarat
Goa	Maharashtra
Pondicherry	Tamil Nadu