

APPENDIX B: SAMPLE DESIGN

The major features of the sample design are described in this appendix. Sample design features include target sample size, sample allocation, sampling frame and listing, choice of domains, sampling stages, stratification, and the calculation of sample weights.

The primary objective of the sample design for MICS Punjab, 2014 was to produce statistically reliable estimates of most of the indicators, at the provincial level, 36 districts and for urban and rural areas. Urban and rural areas in each of the 36 districts were defined as the sampling strata.

Universe

The universe of this Survey consists of all urban and rural areas of Punjab defined as such by Housing Census 2011. The military restricted areas and cantonment have been excluded from the scope of the survey.

Sampling Frame

Pakistan Bureau of Statistics (PBS) has developed its own sampling frame based upon Housing Census 2011. Each city/town/village/deh has been divided into enumeration blocks. Each enumeration block comprised of 200-250 households on the average with well-defined boundaries and maps. The urban area sampling frame was updated during 2013 and the rural area sampling frame consists of a list of villages/deh/blocks compiled during Housing Census 2011. In the updated sampling frame, enumeration blocks, both urban and rural, are considered as Primary Sampling Units (PSUs). The breakdown of enumeration blocks into urban and rural areas of Punjab are as below:-

Province	Urban	Rural	Total
Punjab	22,415	58,063	80,478

Stratification Plan

A. Urban Domain

Large cities:

There are eight Large cities (Lahore, Faisalabad, Rawalpindi, Gujranwala, Multan, Sargodha, Sialkot and Bahawalpur), having population five hundred thousand and above. The Pakistan Bureau of Statistics (PBS) is maintaining a separate frame for SRC/Major urban for the corresponding 8 districts. Therefore these 8 districts have been divided into three strata each, while the other 28 districts have two strata each. Therefore instead of 72 strata (36 districts into 2), 8 more strata for major urban were added and renumbered from 73 to 80. Each of these cities has further been sub-stratified into low, middle and high income groups. The weights are calculated at the cluster level.

Other urban areas:

After excluding the large cities from the eight administrative districts, the remaining urban areas of each administrative district has been taken as a separate independent stratum.

B. Rural Domain

All Rural areas of each administrative district, has been treated as an independent stratum. The sample selection has been undertaken independently in each district and stratum.

Sample Size and Sample Allocation

Keeping in view the key variables and main objectives of the survey, a sample of 41,000 households has been considered appropriate to yield reliable estimates of population parameters within acceptable reliability limits. For the estimation of the sample size, the key indicators used were the underweight prevalence among children age 0-4 years, immunization coverage, antenatal care and literacy rate. The following formula given in the MICS5 methodology was used to estimate the required sample size:-

$$n = \frac{[4(r)(1-r)(deff)]}{[(0.15r)^2(pb)(AveSize)(RR)]}$$

where

- n is the required sample size, expressed as number of households
- 4 is a factor to achieve the 95 percent level of confidence
- r is the predicted or anticipated value of the indicator, expressed in the form of a proportion
- $deff$ is the design effect for the indicator, estimated from a previous survey or using a default value of 2
- $0.15r$ is the margin of error to be tolerated at the 95 percent level of confidence, defined as 15 percent of r (relative margin of error of r)
- pb is the proportion of the total population upon which the indicator, r , is based
- $AveSize$ is the average household size (number of persons per household)
- RR is the predicted response rate

For the calculation, r (underweight prevalence) was taken from the previous MICS i.e., 32.6 percent. The value of $deff$ (design effect) was taken as 2.0 based on estimates from previous surveys, pb (percentage of children age 0-4 years in the total population) was taken as 12.6 percent, $AveSize$ (average household size) was taken as 6.3, and the response rate was assumed to be 90 percent, based on experience from previous surveys. Furthermore the relative margin of error to be tolerated at the 95 percent level of confidence was taken as 15 percent.

By using the above mentioned assumptions, the number of sample households was estimated for each of the 36 districts, and summed up to the total sample for the province. The total sample was then reallocated to all the districts in proportion to the square root of the population estimates of December 2012. The urban and rural allocation in each of the districts was made according to the proportion of urban-rural population, with slight oversampling for the urban stratum to obtain reliable urban estimates.

The number of households selected per cluster for the MICS Punjab, 2014 was determined as 20 households, based on a number of considerations, including the design effect, the budget available, and the time that would be needed per team to complete one cluster. Dividing the total number of households by the number of sample households per cluster, it was calculated that 2050 sample clusters would need to be selected from the province. Therefore, the entire sample of households (SSUs) was drawn from 2050 Primary Sampling Units (PSUs), out of which 774 were urban and 1276 were rural.

The total sample is allocated to 36 districts in proportion to the square root of the population estimates of December 2012. Therefore, each district had different clusters based on its population as of December 2012. Furthermore, in each district, the clusters (primary sampling units) were distributed to the urban and rural domains proportionally to the size of urban and rural populations in that district. The table below shows the allocation of clusters to the sampling strata.

Table SD.1: Allocation of Sample Clusters (Primary Sampling Units) to Sampling Strata

	Estimated Population December 2012 (000')			Number of Clusters		
	Total	Urban	Rural	Total	Urban	Rural
Total	96,676	30,803	65,598	2050	774	1276
Divisions/Districts						
Bahawalpur	10,339	2,271	8,068	200	69	131
Bahawalpur	3,337	912	2,425	66	25	41
Bahawalnagar	2,641	503	2,138	59	19	40
Rahim Yar Khan	4,361	856	3,505	75	25	50
D G Khan	9,120	1,227	7,894	215	62	153
Dera Ghazi Khan	2,308	321	1,987	55	16	39
Layyah	1,542	199	1,344	45	13	32
Muzaffargarh	3,726	483	3,243	70	19	51
Rajanpur	1,544	224	1,320	45	14	31
Faisalabad	12,585	4,301	8,284	241	97	144
Faisalabad City	7,029	3,029	4,000	96	44	52
Chiniot	1,184	318	866	39	15	24
Jhang	2,354	574	1,780	55	20	35
Toba Tek Singh	2,018	380	1,638	51	17	34
Gujranwala	14,651	4,647	10,004	328	125	203
Gujranwala City	4,555	2,323	2,232	77	39	38
Gujrat	2,577	715	1,862	58	22	36
Hafizabad	1,052	287	765	37	14	23
M. B. Din	1,408	214	1,194	43	13	30
Narowal	1,549	189	1,360	45	12	33
Sialkot	3,510	919	2,591	69	26	43
Lahore	16,363	9,176	6,911	274	134	140
Lahore City	9,015	7,397	1,618	108	74	34
Kasur	3,112	710	2,402	64	23	41
Nankana Sahib	1,256	190	1,066	40	12	28
Sheikhupura	2,980	879	1,825	62	25	37
Multan	11,102	2,880	8,222	237	84	153
Multan	4,127	1,742	2,385	73	34	39
Khanewal	2,662	469	2,193	59	18	41
Lodhran	1,554	226	1,328	45	14	31
Vehari	2,759	443	2,316	60	19	41
Sahiwal	6,832	1,274	5,558	163	52	111
Sahiwal	2,301	377	1,924	55	17	38
Okara	2,865	660	2,205	61	22	39
Pakpattan	1,666	237	1,429	47	14	33
Rawalpindi	8,563	3,322	5,241	203	83	120
Rawalpindi City	4,467	2,497	1,970	76	41	35
Attock	1,604	341	1,263	46	16	30
Chakwal	1,330	162	1,168	42	11	31
Jhelum	1,162	322	840	39	15	24
Sargodha	7,121	1,705	5,416	188	67	121
Sargodha	3,266	919	2,347	65	25	40
Bhakkar	1,390	223	1,167	43	13	30
Khushab	1,119	283	836	38	14	24
Mianwali	1,346	280	1,066	42	14	28

Sample Design

A two-stage, stratified cluster sampling approach was used for the selection of the survey sample.

Stage-1: Selection of Primary Sampling Units (PSUs)

Enumeration blocks in urban and rural domains were taken as PSUs. In urban and rural domains sample PSUs from each stratum were selected by probability proportionate to size. The number of households in each PSU from the frame was considered as the measure of size for the urban and rural domains.

Stage-2: Selection of Secondary Sampling Units (SSUs)

Based on actual listing undertaken in respect of each sample PSU by the Field Staff, 20 households were selected from both rural and urban sample areas adopting systematic sampling technique with a random start. Households were considered as secondary sampling units (SSUs) for urban and rural domains. The sample households were selected within each sample PSU with equal probability.

Sampling Frame and Selection of Clusters

The 1998 census frame updated by PBS recently in 2011, during housing census, was used for the selection of clusters. Census enumeration areas were defined as primary sampling units (PSUs), and were selected from each of the sampling strata by using systematic pps (probability proportional to size) sampling procedures, based on the number of households in each enumeration area from the updated frame. The first stage of sampling was thus completed by selecting the required number of enumeration areas from each of the 36 districts, separately for the urban and rural strata.

Listing Activities

For the selection of households i.e., SSUs, a new listing of households was conducted in all the sample enumeration areas prior to the selection of households since the sampling frame was not up-to-date. For this purpose, listing teams were formed who visited all of the selected enumeration areas and listed all households in each enumeration area.

Selection of Households

Lists of households were prepared by the listing teams in the field for each enumeration area selected in the sample. The households were then sequentially numbered from 1 up to the total number of households in each enumeration area, at the Divisional office of the BoS, where the selection of 20 households in each enumeration area was carried out using random systematic selection procedures.

Calculation of Sample Weights

The MICS Punjab, 2014 sample is not self-weighting. Essentially, by proportionally allocating the numbers of households to each of the districts, different sampling fractions were used. For this reason, sample weights were calculated and these were used in the subsequent analyses of the survey data.

The major component of the weight is the reciprocal of the sampling fraction employed in selecting the number of sample households in that particular sampling stratum (h) and PSU (i):

$$W_{hi} = \frac{1}{f_{hi}}$$

The term f_{hi} , the sampling fraction for the sample households in the i -th sample PSU in the h -th stratum, is the product of probabilities of selection at every stage in each sampling stratum:

$$f_{hi} = p_{1hi} \times p_{2hi} \times p_{3hi}$$

where p_{shi} is the probability of selection of the sampling unit at stage s for the i -th sample PSU in the h -th sampling stratum. Based on the sample design, these probabilities were calculated as follows:

$$p_{1hi} = \frac{n_h \times M_{hi}}{M_h},$$

n_h = number of sample PSUs selected in stratum h

M_{hi} = number of households in the 2010 Census frame for the i -th sample PSU in stratum h

M_h = total number of households in the 2010 Census frame for stratum h

p_{2hi} = proportion of the PSU listed the i -th sample PSU stratum h (in the case of PSUs that were segmented); for non-segmented PSUs, $p_{2hi} = 1$

$$p_{3hi} = \frac{20}{M'_{hi}}$$

M'_{hi} = number of households listed in the i -th sample PSU in stratum h

Since the number of households in each enumeration area (PSU) from the 2010 Census frame used for the first stage selection and the updated number of households in the enumeration area from the listing are generally different, individual overall probabilities of selection for households in each sample enumeration area (cluster) were calculated.

A final component in the calculation of sample weights takes into account the level of non-response for the household and individual interviews. The adjustment for household non-response in each stratum is equal to:

$$\frac{1}{RR_h}$$

where RR_h is the response rate for the sample households in stratum h , defined as the proportion of the number of interviewed households in stratum h out of the number of selected households found to be occupied during the fieldwork in stratum h .

Similarly, adjustment for non-response at the individual level (women, and under-5 children) for each stratum is equal to:

$$\frac{1}{RR_h}$$

where RR_h is the response rate for the individual questionnaires in stratum h , defined as the proportion of eligible individuals (women and under-5 children) in the sample households in stratum h who were successfully interviewed.

After the completion of fieldwork, response rates were calculated for each sampling stratum. These were used to adjust the sample weights calculated for each cluster. Response rates in the MICS Punjab, 2014 are shown in Table HH.1 in this report.

The non-response adjustment factors for the individual women, and under-5 questionnaires were applied to the adjusted household weights. Numbers of eligible women, and under-5 children were obtained from the roster of household members in the Household Questionnaire for households where interviews were completed.

The design weights for the households were calculated by multiplying the inverse of the probabilities of selection by the non-response adjustment factor for each enumeration area. These weights were then standardized (or normalized), one purpose of which is to make the weighted sum of the interviewed sample units equal to the unweighted total number of completed interviews at the provincial level. Normalization is achieved by dividing the full sample weights (adjusted for nonresponse) by the average of these weights across all households in the full (provincial) sample. This is performed by multiplying the sample weights by a constant factor equal to the unweighted number of households at the provincial level divided by the weighted total number of households (using the full sample weights adjusted for nonresponse). A similar standardization procedure was followed in obtaining standardized weights for the individual women, and under-5 questionnaires. Adjusted (normalized) weights varied between 0.21 and 6.13 in the 2050 sample enumeration areas (clusters).

Sample weights were appended to all data sets and analyses were performed by weighting households, women or under-5s with these sample weights.