

CHAPTER-2

ORGANISATION OF THE POST ENUMERATION SURVEY

The PES was conducted for the 2001 Census in all the states and union territories after the population enumeration, between April-July 2001.

2.2 The preparation of the PES started with the determination of the sample size, sample design and the schedules for canvassing. The experience gained in the Post Enumeration Check, 1991 was used to improve the methodology as well as the data collection procedure. In 1991, for estimating Type II error, only a 10 percent sub sample of households enumerated in census was selected from each selected block. In PES 2001, all the residential households in the selected enumeration blocks (EBs) were enumerated for estimating the Type I and Total errors and the Type II error was derived as the residual. This effectively reduced the sample size required to estimate levels with the same precision as in 1991. The 1991 PES omission rates and the Percent Relative Standard Errors (PRSE) were considered to estimate the total sample size for the conduct of the PES 2001. The sample size depends not only on the level of disaggregation and the level of reliability required for analysis, but also on resource constraints. The non-sampling error plays an important part in the overall accuracy of the estimates. Keeping all these in view, it was decided that the overall sample size might be limited to a level that is operationally manageable in field and during matching operations. On these considerations, it was decided to fix the sample size of 3,000 EBs at the national level.

2.3 States have been grouped into zones and the estimates are presented at zonal level. At the zonal level, they were found to be reasonably precise in terms of the PRSE. Hence, the estimates have been presented at the national and six zonal levels. The composition of the zones is as below: -

Southern Zone: Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Goa, Pondicherry and Lakshadweep

Eastern Zone: Bihar, West Bengal, Jharkhand, Orissa, Sikkim, and A & N Islands

North Eastern zone: Assam, Arunachal Pradesh, Nagaland, Mizoram, Manipur, Tripura, and Meghalaya

Northern Zone: Jammu & Kashmir, Himachal Pradesh, Punjab Chandigarh, Haryana, Delhi and Rajasthan

Western Zone: Maharashtra, Gujarat, Daman & Diu and D & N Haveli

Central Zone: Uttar Pradesh, Madhya Pradesh, Chhattisgarh and Uttaranchal

Allocation of the sample

2.4 The sample of 3,000 EBs have been distributed among all the States and Union territories in proportion to their projected population of 2001 based on 1991 Census, subject to minimum of 100 EBs for major states (having a population of 10 million and above) and 4 EBs for smaller states/union territories. In case of Jharkhand, Chhattisgarh and Uttaranchal, the sample size were determined on the basis of population of combined Bihar, Madhya Pradesh and Uttar Pradesh and then the share of new states were taken out. For the purpose of studying the content error, only a 20 percent sub-

sample of the EBs selected for estimating the coverage error was used. The sample size at the state level is given below: -

Sample Enumeration Blocks		
Serial Number	State/Union territory*	Number of EBs
1	2	3
	INDIA	3000
1	Jammu & Kashmir	100
2	Himachal Pradesh	52
3	Punjab	100
4	Chandigarh *	7
5	Uttaranchal	34
6	Haryana	100
7	Delhi *	100
8	Rajasthan	139
9	Uttar Pradesh	409
10	Bihar	189
11	Sikkim	4
12	Arunachal Pradesh	9
13	Nagaland	13
14	Manipur	20
15	Mizoram	7
16	Tripura	31
17	Meghalaya	19
18	Assam	100
19	West Bengal	204
20	Jharkhand	70
21	Orissa	100
22	Chhattisgarh	59
23	Madhya Pradesh	148
24	Gujarat	125
25	Daman & Diu *	4
26	Dadra & Nagar Haveli *	4
27	Maharashtra	235
28	Andhra Pradesh	195
29	Karnataka	134
30	Goa	13
31	Lakshadweep *	4
32	Kerala	100
33	Tamil Nadu	159
34	Pondicherry *	9
35	Andaman & Nicobar Islands *	4

Sampling frame and sample selection

2.5 The sampling frames required for the PES were made available to Headquarters of the Office of the Registrar General, India (ORGI) by the concerned Directorates of Census Operations (DCO) in the prescribed proforma, in which the number of enumeration blocks in each charge was mentioned. This information was collected separately for three strata, cities, non-city urban and rural areas.

2.6 For selecting the sample, all the census charges in a state in a particular stratum were arranged in the order of their location code. The required number of EBs, were selected linear systematically after keeping all the strata together. A 20 percent sub-sample of the 3,000 EBs selected above, was taken linear systematically for the survey on content error and all the residential households of these 20 percent or 600 EBs were enumerated completely for this purpose. Due to operational considerations, the enquiry on content error was limited in nature and its scope was restricted to a few items like age, literacy, marital status, economic activity, fertility etc. The sample selection was completely done at the head quarters of the Office of the Registrar General, India, New Delhi and the list of the selected EBs was communicated to the states. It is pertinent to mention in this connection that although 3,000 EBs were selected for conduct of PES, due to earthquake in Gujarat, and other problems in Punjab and West Bengal, the PES could not be conducted in 7 EBs, the break-up being, Gujarat (4), Punjab (2) and West Bengal (1). So, effective number of EBs in which PES was conducted for Census 2001 was 2993.

Data Collection Methodology

2.7 The unit of analysis in the PES is the household member. The target universe included the residential and partly - residential households. The individuals living in the institutional and houseless households constitute only a small proportion of the total population. The composition of institutional households like students' hostels, *Dharamsala*, *Dak bungalow* etc. could have changed during the period after the census. Similarly the houseless households may not be available at the place, where they were enumerated. Hence these two types of households were excluded from the scope of the PES.

2.8 Three main schedules were canvassed in this survey. These are PES Schedule I, PES Schedule IV and PES Schedule VI. The first two schedules relate to coverage error and the last one relates to content error. The basic purposes of canvassing the three schedules were given below: -

1. **Schedule I:** to identify the households, which have been omitted or duplicated; in other words, to determine Type I error
2. **Schedule IV:** to find out persons omitted or duplicated in households which have been enumerated in the census; in other words, to determine Total error
3. **Schedule VI:** to determine content error in selected questions. This schedule was to be canvassed only in a sub-sample of the PES EBs.

2.9 In addition to the three schedules mentioned above, three more schedules, namely, Schedule II, III and V were also used at the time desk matching and field reconciliation. Instructions issued to fill up the different schedules are given in Appendix I.

2.10 In order to facilitate the enumerators to locate the EB properly, a list of 10 percent sub-sample of censused households of the concerned EB was provided to the enumerator before going to the field for re-listing and re-canvassing of the PES Schedules. In addition to that, layout map of the concerned EB together with those of the adjoining EBs were provided to him at that time. Efforts were made to make the PES as independent as possible, operationally, from the census. For that, following measures were adopted.

- (i) Before the commencement of the PES in a state/ut, it was ensured that all the census enumeration records of that state/ut were brought to the concerned DCO.
- (ii) The list of the selected EBs was not disclosed to the DCOs till the population enumeration was over.
- (iii) To ensure independence of the PES from Census the Abridged House List and Census Household Schedule were not shown to the enumerator at the Independent initial stage

Stages of work of the PES

2.11 There were three distinct stages of operation for conducting the PES as described below:

- (i). *Independent listing and independent canvassing of the PES Schedules I, IV and VI (wherever applicable): - a field operation*

2.12 The Schedule I had been filled for all the census houses, whether residential or non-residential, in each of the selected EBs. In other words, the re-listing was done in each of the selected EBs. These lists were compared with the respective Abridged House list in the DCO, using certain criteria for matching households and wherever necessary, field reconciliation was done by re-visiting that house to identify the omitted/duplicated households. The members of such households were then listed along with certain particulars.

2.13 While filling Schedule I, wherever the PES enumerator came across any residential house/ household, he/she had to fill Schedule IV immediately thereafter, for particulars of the persons in that household. These particulars were compared with the corresponding census household schedule for each individual in that household. For any omission or duplication of any member in the household, field reconciliation was done to sort out the discrepancy.

2.14 As mentioned earlier, for the survey on content errors, a 20 percent sub-sample of the sample EBs for estimating the coverage error was considered. So, if the EB in which the Schedules I and IV were canvassed is in the sub-sample, then Schedule VI was also canvassed immediately after canvassing the Schedules I and IV.

2.15 After the re-listing and re-canvassing of the PES Schedules, these were brought to the office of the DCO for desk matching. Before the desk match started, the AHL and the Census Household Schedules in respect of the selected EBs for the PES were kept separately so that at the time of matching, these were readily available.

(ii). Desk matching of particulars collected in the PES and the corresponding Census records of an EB: - a desk job

2.16 In the first step of the desk matching exercise, PES Schedule I, Census Household Schedule and the corresponding AHL of the same EB were taken out for matching the respective entries. Once this was assured, then one by one, the entries from the PES Schedule I were matched with those in the AHL. Similarly, the entries in the PES Schedules IV and VI were matched with the corresponding entries of the Census Household Schedule. The matched and non-matched cases were identified on the schedules in the appropriate columns and the Schedules II, III and V were filled up and those records were identified for the purpose of field reconciliation.

(iii). Field reconciliation: - a field operation

2.17 In this stage of the PES, households or individuals were revisited and reconciliation was done regarding the entries, which were not matched between the PES and the relevant Census records.

2.18 PES estimation methodology is based on the assumption of independence between the Census and the PES operations. To maintain independence, different sets of persons were engaged for re-listing and re-enumeration operations on the one hand and the desk match and field reconciliation operations on the other.

Organisational aspects

2.19 On the basis of the time required for filling up the schedules, about ten days per EB were considered necessary for the work of listing, desk matching and field reconciliation. As the survey had to be conducted as close to census as possible, it was felt that an enumerator should be allotted the work of only one EB.

2.20 As mentioned above, the PES estimation methodology is based on the assumption of independence between the actual census and PES operations. To maintain independence, in addition to the officers and staff of the DCOs, the officers and staff of the Directorate of Economics and Statistics (DES) of the respective State Governments and other government organizations were involved in most of the States and union territories. The enumerators were of the ranks of Assistant Compilers, Compilers, Statistical Assistants, etc. Their work were supervised by the officers of the rank of Senior Statistical Assistants, Investigators, etc. Overall responsibility of conducting the PES in a State/Ut was assigned to a senior officer of the DCO, especially appointed as a nodal officer for that purpose. Wherever fieldwork was done by the DES, one nodal officer from the State Government was also appointed so that both the nodal officers could have close liaison in accomplishing the work in the specified time. Field

inspections were also undertaken by the Directors of Census Operations and the senior officers of the Office of the Registrar General, India (ORGI). In order to ensure application of uniform yardstick in canvassing and thereby arrive accuracy of results, the officers of the DCOs and the State Governments, appointed as nodal officers, were imparted training on the concepts and methodology of conducting the PES by the officers of the ORGI. They, in turn, imparted similar training to the enumerators and supervisors under their control.

Estimation Procedure

2.21 The number of persons enumerated in census, omitted in census and duplicated in census has been estimated at the state level separately for rural and urban areas by sex. Since census and PES were conducted by different sets of enumerators, and the two operations were independent, dual estimation procedure has been used to estimate the persons, who were omitted by both census and PES. This procedure is described below:

2.22 The population eligible for enumeration in census can be categorized as those enumerated in Census and the PES, those enumerated in PES only, those enumerated in Census only and those not enumerated in either PES or Census. This can be presented in tabular form as below:

Enumerated status in			
PES	Census		
	Enumerated	Omitted	Total
Enumerated	N_{11}	N_{12}	$N_{1.}$
Omitted	N_{21}	N_{22}	$N_{2.}$
Total	$N_{.1}$	$N_{.2}$	$N_{.}$

2.23 Three of the components of the total population, N_{11} , N_{21} and N_{12} are available from the survey and census data. The number of persons omitted by both the PES and Census is not available. Under the assumption of independence of the PES and Census, the unknown quantity N_{22} can be estimated as

$$N_{22} = \frac{(N_{12} \times N_{21})}{N_{11}}$$

2.24 The dual-system estimation procedure is based on the case-by-case matching of two different and independent sources describing the same event. The dual estimation procedure was applied at the zonal level by sex separately for rural and urban areas for estimating aggregate number of persons omitted, and also in case of categories like literacy, marital status, age, etc. The net omission has been obtained by subtracting the number of persons who have been duplicated from the number of persons omitted. The net omission rate presented in this report gives the ratio of the number of omitted persons (net of duplication) per 1000 persons enumerated in census.

2.25 Those enumerated in census have been classified as:

1. Enumerated correctly: These persons are enumerated in the correct place only, i.e. in places where they were eligible for enumeration as per census instructions.
2. Enumerated only once but at wrong place: These persons have been enumerated at a place other than the place in which they were eligible for enumeration.
3. Duplicated: These persons were enumerated in more than one place.

2.26 The above classifications were made on the basis of the questions canvassed in PES. The persons who were enumerated once but at wrong place were considered as enumerated in census for purposes of calculating the omission rates. Though at the national level this is correct, it can lead to distortion of results at the lower level. However, these distortions are unlikely to alter the results significantly in view of low proportion of population in this category. The standard errors (SE) of the omission rates have been estimated for total error. The formulae used for estimation of variances and SE is given in Appendix III. The definition and usage of indices of content error are explained in the chapter on content error.

2.27 The PRSEs have been estimated and presented for the total error by sex and residence at the national and zonal levels only. The PRSEs have not been estimated for the error rates based on dual estimation. As the dual estimation procedure has been applied only at the zonal level and not at the enumeration block level or state level, the standard errors for these estimates could not be estimated in the same way as that of the direct estimates. Standard errors for the estimates based on dual estimation are likely to be higher than those for the direct estimates. In other words, the PRSEs presented are slight under estimates. In case of the error rates for categories like literacy, marital status, age, etc., the PRSEs have not been worked out. The estimated coverage error based on the dual estimation procedures is 23 persons per thousand at national level. The zonal level omission rates are in the range 8 to 49 persons per thousand persons