

## **Draft Report**

# **POST ENUMERATION SURVEY 2001**

Raja B. M. Korale  
UNFPA PES Consultant

## **Executive Summary**

The aim of a post enumeration survey is to assess the completeness in the census enumeration and also the quality of the answers given to the questions asked in the population census. An independent verification of the census enumeration through a PES on a sample basis can provide an estimate of the extent of under enumeration or over enumeration that occurred at the census.

The post enumeration studies have shown that the undercount is different in different geographic areas and for different demographic and socio-economic groups. It would be highly desirable though infeasible, to know how many people the census missed and from what locations, then they could be added to where they belong to improve the undercounts. The principal object of the PES 2001 was to estimate the undercount rates so that the estimates could be used to adjust as required, the undercount of the enumerated population at the census .

The censuses of population have been conducted decennially in Nepal commencing from 1911. The Census of Population 2001 is the 10<sup>th</sup> Census undertaken of this series of population censuses taken in the modern period. The conduct of a population census in a mountainous country such as Nepal where logistical problems are daunting will remain a major statistical undertaking. As it should be expected there will be under enumeration and over enumeration at these censuses. A post enumeration survey (PES) helps to evaluate the census operations and ascertain the degree of undercount. The PES undertaken in 1991 had shown that the under enumeration at the census was as high as 11%, and due to certain deficiencies in the PES this under enumeration has been questioned at the time.

The Census of Population 2001 was taken with 22nd June as the reference day. According to the preliminary results, the total enumerated population was 23,214,681. The Census 2001 was conducted according to the de-jure approach. According to this method all persons who usually resided at the address of location were counted as having resided at their usual place of residence on the reference day. The population was enumerated using trained enumerators and the persons who should be counted were defined in the instructions prepared for the guidance of census enumerators and supervisors. According to the census law, all persons who usually resided in the land of Nepal should be counted, and it is an offence not to disclose information to the questions canvassed by a census official. Although penalties are specified in census legislation, the CBS has not prosecuted any persons for failure to have themselves enumerated. The Census of Population conducted in June 1991 enumerated the total population as 18.49

million. Thus the inter-censal increase in population has amounted to 5.3 million and the inter-censal population growth rate has amounted to 2.27% per year.

### **Operational Strategy**

The PES 2001 was planned as an independent intensive re-interviews of all households in the sampled enumeration areas. The sample was restricted to a manageable size as mentioned elsewhere. A single stage stratified sampling design was adopted for the household enumeration sampling 7900 households and a two stage stratified design was used for the individual questionnaire. The Dual System Estimation method was adopted for the survey design.

The fieldwork on the current post enumeration survey was conducted during the period between October - November 2001 three to four months after the conclusion of the census enumeration. This delay has contributed to issues on the mobility of persons between the census and the survey. Ideally the post enumeration survey should have been conducted within a month after the conduct of the census to avoid the issues of population mobility and errors that arise due to memory lapses.

### **Results**

According to the PES the undercount rate of the total population of Nepal at the Census of Population 2001 was 6.44%. This is the gross undercount rate and it does not take into account the over-count of the population due to erroneous enumeration such as double counting etc. Therefore the survey estimate is the upper bound of the under-enumeration. The over-count rate could not be computed as the CBS had found it difficult to match the individual population sample and conduct the re-interviews due to staffing and security problems.

The omission rates for the regions and the precision of the estimates are set out in Table 4 below. As expected the undercount rates are higher in the urban areas but they did not have much impact on the overall rate as the relative size of the urban sector amounts to only 14% of the total population.

Table 4  
National and Regional Undercount Rates and the Precision of the Estimates

Sector	Census 2001 Enumerated Population	Coverage Rate	Omission Rate(r)	Standard Error Se(r)	Coefficient of variation cv(r)
Nepal	23,214,681	0.9355	0.0644	0.0069	0.108
Rural Sector	19,911,181	0.946	0.0539	0.0046	0.0853
Tarai	9,744,810	0.952	0.0474	0.0059	0.126
Hill and Mountain	10,166,371	0.939	0.061	0.0071	0.117
<b>Urban Sector</b>	3,303,500	0.877	0.1222	0.034	0.270
Kathmandu	696,852	0.890	0.109	0.067	0.611
Other Urban	2,606,648	0.874	0.125	0.039	0.311

### **Adjustment of Census Estimates**

After each census CBS has produced population estimates and projections for Nepal and sub-national areas. In the construction of these estimates the enumerated population at the census had been taken as the base population. In the estimates the undercount at the census had not been taken into account in adjusting the base year population. The inter-censal population estimates and projections will be more complete and reliable if the population adjusted for the undercount in the census is taken as the base population.

### **Conclusion/Suggestions**

1. The PES 2001 was an advance over the PES conducted in 1991, most importantly in undertaking field work in all the sample enumeration areas which was not achieved in both previous PES surveys conducted in 1981 and 1991. In addition fieldwork was undertaken by centrally trained staff and matching operations and re-interviews were conducted in a more organized and a systematic manner. However, it should be noted that the sample size had to be restricted in view of both funding and manpower constraints. The sample size chosen was barely adequate to realize statistically reliable estimates at aggregate and sectoral levels. Yet the CBS found it hard pressed to provide staff inputs to meet the planned outputs from the survey. The survey dimensions should be determined taking the capacity of the CBS to successfully meet the target outputs.
2. The plans to conduct a full blown PES would bring some pressure on the population planning and implementing staff that the census results would be evaluated and this would prove useful in improving the coverage and quality of census data.
3. In addition to undercounting, the problem of miscounting that results in over-enumeration also appears to be serious issue in census data. The paucity of staff had

prevented the CBS from quickly processing Form 2 and following up on dual enumerations and miscounted persons in the field to use the data for estimation and analysis and preparation of over-enumeration estimates.

#### **4. Census Cartography**

The lack of cartographic materials for census mapping is undoubtedly the main deficiency that had affected completeness and quality of census and survey data. The CBS should establish a Census Cartographic Unit with basic equipment and facilities and 2-3 statisticians should be trained in census mapping, as soon as possible, securing technical assistance funds. The field staff deployed on surveys should be detailed to prepare a sketch map of the sampled enumeration area in all sample surveys to reduce under-enumeration and miscounting.

#### **5. Sample Survey Unit**

The CBS should strengthen the Household Survey Section to support the technical work on household and establishment based surveys. The paucity of trained staff prevents the CBS from benefiting from the statistical surveys currently being undertaken and from the work presently being done utilizing the services of expatriate staff recruited for different project tasks. CBS should arrange to train 2 statisticians in survey design and survey sampling under technical assistance projects.

### **1.0 Introduction**

The census results are used for diverse purposes, such as updating population estimates, developing and updating sampling frames, correcting and updating population registers, delimitation of electoral districts, provision of government budgetary allocations to local authorities, etc. These many uses point to the need for an objective method for assessing, coverage and content errors as an important step in concluding a census operation. The quality assurance checks built into census operations cannot ensure an evaluation of the qualitative and quantitative accuracy of census data.

Although several methods have been developed to evaluate census of population data, for many developing countries the PES serves as the most appropriate method due to a paucity of data required to effectively use such techniques as demographic analysis. Incompleteness of vital registration systems and infrequency with which demographic surveys are conducted limits the use of other methods of census evaluation. Thus, a number of countries have mainly used post enumeration surveys to evaluate census results. Thus, PES's are now extensively used in census evaluation. In summary the PES is a well designed, independent, household based sample survey that replicates a census. The survey and the census results are then compared through a process of matching records and the results of the comparisons are used to measure the coverage errors and often the content errors of the census. The estimates of net coverage, the number of persons omitted in the census, the number erroneously enumerated and content error rates of specific questions are typical products of a PES. Depending on the scope and dimensions of the PES this estimate can be broken down further so that reliable estimates of undercount or over count can be obtained for the entire census, for geographic areas of

interests as well as for demographic characteristics such as age, race, sex, marital status, etc.

A census of population could make two kinds of counting errors about persons enumerated viz.;

- gross omissions or failure to count someone where he or she belongs; and
- erroneous enumerations that arise through reporting someone fictitious or real, where he or she does not belong.

Gross omissions reduce the count, while erroneous enumerations inflate the count. The same individual could contribute to both kinds of errors. In a case where a person's address is recorded incorrectly, he can be a gross omission at his correct address and an erroneous enumeration at his incorrect address. The two kinds of errors cancel to some extent, but overall, the census misses more people due to net undercount. The post enumeration studies have shown that the undercount is different in different geographic areas and for different demographic and socio-economic groups.

### **1.1 Objective**

The primary objective of the Post Enumeration Survey 2001 was to measure the extent of coverage or the level of undercount or over count in the National Census of Population 2001 conducted by the Central Bureau of Statistics. While the survey focused on the gross under count or over count of the total population, the survey would also attempt to measure the level of coverage of selected variables and topics. In addition to the undercount rate at the national level, the assessment of coverage in selected geographic groupings would be attempted by designing the survey to produce domain level estimates. The upgrading of survey taking capability of CBS by preparing survey and sampling designs, survey instruments and operational procedures and training staff for the Bureau to plan and design post enumeration surveys was also an important objective.

### **1.2 Reasons for Miscount**

The enumeration of 23 million people in a mountainous country such as Nepal where the logistical and other problems are daunting is a major statistical undertaking. In such a large and complex undertaking, it was inevitable that some persons would not be enumerated for a number of reasons and some would be enumerated more than once. Experience shows that those groups who were mobile and difficult to enumerate were more likely to be missed in the enumeration. There are many reasons that cause under-enumeration:

- persons shifting from one house to another at the time of the census;
- household members being away temporarily on work, studies, schooling, hospitalization, etc.;
- several households occupying the same premises/address;
- persons with no fixed abode;
- persons deliberately avoiding the census (for fear that information given will be used against them);
- new born babies being over looked;
- dwelling entirely missed by enumerators and occupied dwelling misclassified as vacant;

- failure of census enumeration procedures.

Over enumeration also results from several causes and mostly due to double counting. These include the multiple enumeration of persons with more than one residence; persons living away from home while working or studying and enumerated at home and at the place of stay; persons shifting from one house to another around the time of the census; erroneous enumeration of diseased persons; migrants who had left the country permanently and babies born after the census night.

### 1.3 Need for Undercount Estimates

The census of population statistics are used for a wide variety of purposes and it is essential to have accurate estimates as feasible from the census. Further the census data are used as baseline data to derive inter-censal estimates of population. Thus, the census estimates are used for:

- denominators in many socio-economic indices, e.g. birth rate, death rate, per capita income, etc.;
- population frames for various surveys;
- planning, policy making and administration;
- demographic and socio-economic research;
- allocation of electoral divisions and demarcation of parliamentary electorates.
- allocation of public funds from the government budget for different socio-economic development activities

### 1.4 Previous Post Enumeration Surveys

A PES was conducted for the first time in Nepal in 1981 to estimate the under count of the Census of Population 1981. Although, the PES was planned to be undertaken in 67 wards, the survey results had been produced only in respect of purposively selected six rural wards. The under-coverage or the omission rate was reported as 14.1% with a standard error of 0.031%, and the coefficient of variation was estimated as 21.98%. The estimate of under-coverage was based on one way matching only ie PES records were matched with census records. The reported omission rate was considered as an over-estimate in studies undertaken at the time. The survey as implemented was more a pilot study that was restricted to a survey of purposively selected six rural wards. The under coverage rates would therefore be applicable only to the sampled wards and not to (the sample was too small anyway) the country as a whole.

A well-designed PES had been planned to check the coverage errors in the Census of Population 1991. The sample had comprised 182 rural wards and 8 urban wards with an estimated 116,000 persons in 21,000 households. The sampling design had involved the stratification of the country to Terai, Hill and Mountain regions and urban and rural stratifications. A two stage stratified systematic sampling procedure had been adopted in the selection of the sample with districts as the primary sampling units and wards which were treated as the 2<sup>nd</sup> stage units. Out of the 182 rural wards sampled for the survey, fieldwork had been conducted only in 96 rural wards. In the Terai and Hill regions the survey was not conducted in 62 of the 158 sampled wards. The Mountain region was completely excluded; fieldwork could not be undertaken in any of the 24 sampled wards.

The original sample had been selected following systematic sampling procedure, although it could not be surveyed fully. The final purposively selected sample contained 12,295 households in the rural sample and 2,922 households in the urban sample a total of 15,217 households. In these households 87,134 persons had been recorded.

The PES 1991 had estimated the omission rate as 11.05% and the coefficient of variation was reported as 7.41%. Here again the reported omission rate was treated as an over-estimate. The evaluation of the coverage error using demographic analysis techniques had shown that it was of the order of 3% - 5%. The PES has not reported the design effect but when it was computed it was found to be as high as 59. An examination of the operational and estimation procedures adopted in the PES showed that the reported omission rates could in fact be overestimates.

There a couple of important issues that should be noted here in view of the implications they have had on the PES. Firstly, a failure to enumerate any sampled wards could increase the cluster effect (as the excluded wards would in fact be from less accessible locations and would be dissimilar to the wards that were enumerated) apart from raising the amount of bias in the estimates. More importantly, the PES 1991 was also an ambitious project in relation to the capacity of the CBS to successfully implement a large survey, immediately after fieldwork on a complex activity such as a census of population while census processing was ongoing. Thus, a major consideration in deciding on the scale of the PES was the capacity of the CBS to successfully undertake the PES as planned.

## **2.0 PES Design**

The survey objectives described earlier have by and large determined the survey design. The survey had to be nationwide and provide statistically reliable estimates for the main geographic stratifications. However, practical considerations made it necessary to use a truncated frame that excluded areas that were considered not safe to undertake fieldwork for security considerations. The previous experience of undertaking PES also suggested that the survey should be kept to a modest sample size in the context of difficulties encountered in carrying out any large scale survey immediately after a population census while census processing was ongoing. In some countries PES has adopted the sample design and infrastructure generally used in current population surveys or household survey programmes. CBS does not currently undertake a regular household survey programme. The CBS was keen to adopt the strategy adopted in PES 1991 of dual system estimation method (DSE) and also restrict the household enumeration to a listing operation to reduce the workload involved in fieldwork.

### **2.1 Scope of the Survey**

For practical reasons the scope of the PES 2001 was restricted to private dwellings, both permanent and temporary including one person households. Households and persons who had no fixed abode were thus excluded. Accordingly, PES 2001 was based on a stratified sample of about 8,500 households residing in private dwellings both temporary and private in the sampled wards which were totally enumerated. For the same reasons non-private dwellings including institutional households, such as boarding

houses, army barracks, hospitals, prisons were also excluded as persons in these institutional households cannot be treated as usual residents in them. Further, it will be practically difficult to conduct a PES and collect comprehensive information from such non-private households as persons who resided at these institutions may have changed their residential status after the census day. As is usually done in household based surveys the following population subgroups were excluded from the PES:

- persons living in institutional households or non private dwellings e.g. hostels, hospitals, prisons and army barracks;
- persons who had no fixed abode;
- persons who died after census night;
- babies born after census night; and
- foreigners who visited the country.

As far as the persons who had no fixed abode were concerned, the difficulties in enumeration and matching also meant that homeless people were beyond the scope of the PES. However, as the process was finally applied they would have received the same adjustments in estimation of under count as persons occupying residential dwellings. As stated above, it was also necessary to exclude some disturbed areas of the country where it was considered unsafe to conduct fieldwork on the PES. Further more, it had not been possible to make a complete count in the Census 2001 in these areas for security considerations. However it should be noted that these excluded areas from the PES may have been problem areas at the census where higher levels of omissions and under counts could have resulted. This means that the adjustments to these categories of excluded population will have to be determined from the sample of persons in private households in residential housing units where stable conditions had prevailed.

## **2.2 Operational Independence**

The PES should be conducted independently of the census as feasible to maintain the integrity of the PES. The formula that is used in estimating the coverage and omission rates depends on this independence between the PES and census. In order to achieve this:

- i. PES sample is drawn from a sampling frame other than that of the census;
- ii. census field staff are not usually deployed in PES field work;
- iii. the information on enumeration areas sampled for the PES should not be disclosed to the census enumeration staff.
- iv. PES fieldwork should commence after the completion of the census enumeration to prevent an overlap.

In conducting the PES, several safeguards were made to achieve these criteria to the extent feasible. Sampling frame used for the PES was not based on the Census 2001, but it comprised an updated version of the frame prepared from the Census of Population 1991 for rural areas and a list based on the pre-listing operation conducted for the Census 2001 for the urban areas. Mainly primary school teachers were deployed in the Census of Population 2001 enumeration work and CBS staff were not deployed on enumeration duties. The staff deployed for the PES was CBS staff from head quarters and from the District Statistical Offices who had not functioned as census enumerators. Most importantly the PES field staff were trained in Katmandu for 3 days by the senior staff of

the CBS and this issue was discussed and explained to them. The matching of records was done at head quarters by CBS staff who were not involved in census enumeration or processing operations. These procedures were considered adequate to assume the independence of operations between the census and the PES. In spite of these precautions to maintain independence between the PES and the census, yet PES results may have a co-relation bias arising from factors that contributed to persons being missed in the census, which may also cause them to be missed in the PES.

### 2.3 Dual System Enumeration

Because there are two types of coverage errors, namely omissions or missed persons, and erroneous inclusions two samples are selected to evaluate census coverage. In the dual system of enumeration (DSE) the two samples are described as the population sample (P –Sample) which is used to estimate omissions) and the enumeration sample (E-Sample which is used to estimate erroneous inclusions). P-sample consists of persons drawn from a sampling frame covering the entire country. The persons in the P-Sample are matched back to the Census to determine whether they were counted or missed. The E-Sample is drawn from the Census and consists of persons living in a sample of housing units enumerated in the census. The E-Sample persons are checked to determine whether they were correctly counted in the census, or whether they were erroneously included. Erroneous enumerations include dual enumerations, fictitious names, persons born after the census day and persons who died before the census day.

#### Two Types of Errors

	Omissions	Erroneous Inclusions
	<u>P Sample</u>	<u>E Sample</u>
Estimates	Omissions	Erroneous Inclusions
Universe	All housing units	Census housing unit
PSU's	Wards/Sub-Wards	Wards /Sub-Wards

The sample design for PES had to provide for some special considerations. Since the census had used the ward/sub-ward as the enumeration area, a listing of housing units in a ward in the PES could be matched to a listing of a ward from the census. If the entire ward is sampled for the PES, then all housing units and households will have to be listed. Another advantage of taking the ward as the sampling unit was the ability to take the census listing and match back to the PES listing. This allows the E sample to be drawn as the set of wards that coincides with the P sample.

### 2.4 Questionnaire Design

The methodology that was adopted for the PES design had determined the number of questionnaires as well as the content of questionnaires that were canvassed in the PES. Post enumeration surveys conducted by most countries have adopted a single post enumeration survey to derive the estimates. In the USA however, PES conducted in 1980, 1990 and 2000 had adopted the DSE method with the two-way match approach to improve the quality of estimates. In the PES conducted in 1991, the CBS had adopted the

DSE method and had decided to use the same method in PES 2001. This method requires the canvassing of two questionnaires, one from the sampled enumeration areas selected for the post enumeration survey and a second questionnaire from a sample of households enumerated in the Census for the enumeration area sample for the erroneous enumeration survey. CBS had adopted the simplified procedure of canvassing a listing sheet from households in the sampled wards with information needed for matching, in the household listing stage itself and the second questionnaire was canvassed from a sub-sample of households in the same wards. The PES questionnaires were designed to canvass information relating to persons who resided in private households located within the wards/enumeration areas sampled for the PES.

#### **2.41 Household Listing Questionnaire**

The PES questionnaire should be brief. The primary objective of the PES was to ascertain coverage errors in the census for a manageable selected number of characteristics. This is especially important in situations where the population is not highly literate and identification documents, such as personal identification cards, birth certificates, marriage certificates are not available with a majority of the population to assist the enumerators during the PES. The Listing Sheet (Form 1) was designed to cover the following topics:

- basic demographic information for the household head and the eldest member of the household present (name, sex, age, );
- whether the household was visited and household members were counted in the census;
- information on out migrants/ deceased after census day;
- other questions designed to meet the needs of matching process and in the calculation of the under count estimates.

In filling Form 1, the enumerator was required to record the identification and other details of the sampled ward. The questionnaire provided for recording the serial number of the household; ownership status of dwelling; number of persons both male and female enumerated at the census and the number of usual residents at the time of the PES interview; names of head of household and eldest member of household and their age, sex, residential status, and any changes in household composition due to in or out migration. Provision was also made to record information needed for control purposes including dates of enumeration etc.

#### **2.42 Individual Questionnaire**

In order to identify the over count, this questionnaire should attempt to obtain a complete list of all persons who should have been included in the census questionnaire. Eliciting information on persons who were absent at the time of the PES, but who resided when the census was taken could provide the requisite information. Personal information canvassed in the Individual Questionnaire (for the E-sample) Form 2 from the sampled population included selected items which were in the census form and other information including name, surname, sex, age, and usual place of residence and status of enumeration at the census. The collection of information on current usual place of residence, place of residence during the period of the census, place of enumeration and information on any other addresses where the person might have been included on any

other census forms enables the matching of the person at any of the addresses and identification of multiple counts for a particular person. Information on persons who had moved since census day is essential in the estimation process and this question was included.

Several drafts of the two questionnaires were prepared and the drafts were discussed with the CBS officials and revised versions were prepared. The final version of draft questionnaires are attached (See Annex 2). In addition, Notes for Preparation of the Enumerators Manual were also prepared and the document is attached as Annex 3.

### **3.0 Sampling Design**

The sample design used in the PES 1991 was reviewed, and several design modifications were decided upon. The number of stratifications was reduced; the restricted sample size was one consideration for this decision. In order to produce separate estimates for urban and rural sectors, they were treated as two domains. The rural sector was divided into two strata, Terai Region as one stratum, and Hill and Mountain Regions taken together as the second stratum. Similarly, the urban sector was also divided into two strata with Kathmandu Metropolitan area as one stratum and all other urban areas taken together as the second stratum. A single stage probability sample design with the wards and sub-wards as primary sampling units (PSU)'s was adopted for the listing operation ( Form 1) and a two stage design was adopted for the Individual Questionnaire –Form 2. The sampling strategy adopted is described in the paragraphs that follow.

### **3.1 Sampling Frame**

The compilation of the sampling frame from the pre-listing operation carried out for the census 2001 for the rural sector which accounted for over 85% of the population was on-going at the time, and would have taken several months to finalize. The frame for the urban sector from census 2001 was compiled and made available. Thus, it was necessary to fall back on the sampling frame derived from Census 1991 although it was dated. This frame had not been updated, and about 10 days time was taken to check and assemble the frame with the assistance of the CBS officials. The VDC's that were upgraded and attached to urban areas were removed from the rural sector. Two districts that were incompletely enumerated due to unstable conditions in the areas were excised from the frame. The conduct of fieldwork in those areas was considered as difficult for security considerations.

In the Census 1991 sampling frame, the basic unit of enumeration was the ward, with identifiers and the measure of size in households and population totals and breakdowns by males and females. The measure of size should be reasonably current and accurate. The completeness of the sampling frame also depended on the stability of the frame units. As regards the information on households/ population recorded in the frame might have been valid at the time the data was collected. The stability of this data was weak even under normal circumstances due to in or out migration of persons, subject to a number of socio-economic factors. The wards in urban areas and those close to development schemes in the rural areas were the units most likely to suffer from the

changes even over a short period of time. The frame which was 10 years old and the fact that it had not been updated was also a consideration in deciding on the scale of the PES.

An important property of the sampling frame is that the frame should consist of well defined units. Although, the boundaries of wards and particularly sub-wards were not supported by maps and well delineated, yet the village level officials were expected to know their ward boundaries and land marks and the enumeration staff could seek their assistance to correctly identify the boundaries. The wards in rural areas and sub-wards in urban areas were chosen as the sampling units, as the Census 2001 had used them as enumeration areas. The Census 1991 sampling frame was used for sampling in rural strata and the frame from Census 2001 was used for urban strata.

### **3.2 Sample Size Determination**

The discussions with the CBS officials indicated that the sample size should be restricted to a “manageable size”. Both staffing and funding considerations were cited. The note prepared by the UNFPA National Consultant had restricted the sample to a modest 5,000 households. Most importantly, the previous experience of undertaking PES in 1981 and again in 1991 pointed to the difficulty of successfully operationalizing and completing any sizable PES immediately after a population census operation.

CBS had completed the census under difficult circumstances and the enumeration in two districts had been practically abandoned as stated earlier. A truncated frame excluding these districts was used in designing the PES. The CBS senior officials felt that the omission rate could be of the order of 5%. Because of the issues mentioned earlier, PRSE of even 5-7 was acceptable under the circumstances. Then, there was also the issue of the completeness and reliability of the sampling frame. The sampling frame as already discussed was 10 years old and it had not been updated, except for the revisions effected to transfer the areas that were declared urban from the rural to the urban frame.

The design effect computed from the estimates available from the PES conducted in 1981 amounted 15.18. What was most disconcerting was the fact that the design effect estimated from the PES 1991 was much higher, the estimate being over 50. The difficulties of operationalizing the planned design, the altogether exclusion of 24 sampled wards in the Hill Region, and exclusion of wards selectively from the other sampled strata, and a departure from the standard design of canvassing the Individual Questionnaire based on the census sample to that enumerated in the survey had contributed in raising the design effect. Therefore, PES 1991 was not a good indicator of the design effect that would result if the survey had been executed as planned.

The sample size required for assumed omission rates and design effects, for selected values of PRSE (percentage relative standard error = standard deviation divided by the mean x100 ) is shown in Table 1. Assuming that the undercount rate was about 5% then it would be possible to estimate the omission rate with a PRSE of 6% with a sample of 45,000 provided the design effect lies between 7.5 to 10. As stated earlier, CBS lacked capacity to raise the sample and implement a large survey operation at a time when they were required to give priority to census processing operations. Accordingly, it

was decided to limit the sample size to about 7,500 households based on the 1991 census frame for the rural sector and 2001 census frame for the urban sector. In terms of a updated current frame, the sample size would amount about 8,500 sampled households or a sampled population of over 45,000.

### 3.3 Sample Allocation

The sample was rather small to produce sub-national level estimates with low values of coefficients of variation, and this position understood. Within the rural sector, sample could be allocated to Terai region and Hill and Mountain regions in terms of population shares.

It was known that the omission rates were higher in urban areas than in the rural areas. Since it was necessary to have omission rates for the urban and rural sectors separately, it was decided to allocate about 2500 households or 15000 persons to the urban sector. That would allow the omission rate to be estimated with a PRSE of about 12. The final allocation of the sample is shown in table below.

**Table 1: Sample Allocation**

	Stratum	Wards	Households
1	Terai Region	16	2546
2	Hill and Mountain Region	30	2741
3	Kathmandu	4	1370
4	Other urban areas	4	1136
	Total	54	7793

### 3.4 Selection of Sample Wards/Sub-Wards

The frame was arranged so that all the wards/ sub-wards within a stratum were listed in the order of district, VDC, ward/sub-ward with their identification codes and the number households in the ward. The number of households in the ward/sub-ward  $M_{hi}$  was used as the measure of size (MOS) in the probability proportional to size method of selection of first stage units. The method of circular systematic sampling with probability of inclusion of a ward/sub-ward proportional to its size (CSSPPS) method was used to select the sample of wards from each stratum.

The total number of wards in a stratum  $h$  will be denoted by  $N_h$  and the measure of size of the  $i$  th ward by  $M_{hi}$  and the number of wards to be selected from the stratum by the CSSPPS method as a sample will be denoted by  $a_h$ . In this method, the total number of households  $M_h$  in each stratum  $h$  was computed by summing up the measure of size  $M_{hi}$  of each ward  $i$ . The sampling interval  $I$  was computed by dividing  $M_h$  by the number of wards to be sampled  $a_h$  for each stratum, thus

$$I = M_h / a_h, \quad \text{rounded off to the nearest integer.}$$

A random number  $R_h$  that falls between 0 and  $M_h$  was then selected using the random number generator in the Excel programme. The sequence of  $a_h$  selector numbers

were generated by the addition of  $I$  to the previous number selected. If the total exceeds  $M_h$ , then  $M_h$  was subtracted from the total to derive the number.

Let  $R_{h1} = R_h$  and for  $j = 2, 3, \dots, a_h$ ,  $R_{hj} = R_{hj-1} + I$ , if this does not exceed  $M_h$ ;  $R_{hj} = R_{hj-1} + I - M_h$  otherwise.

Accordingly, the selector numbers will be of the form

$R_h, (R_h + I), (R_h + 2I), (R_h + 3I), (R_h + 4I), \dots, (R_h + (j-1)I), (R_h + (a_h-1)I)$ .  
when  $R_{hj} = R_{hj-1} + I$ , does not exceed  $M_h$ . The expressions should be replaced with the terms  $R_{hj} = R_{hj-1} + I - M_h$  when  $R_{hj}$  exceeds  $M_h$ .

Selection of wards in the first stage by CSSPPS was done using Excel. The details of the samples selected for Terai, Hill and Mountain regions, Kathmandu Metropolitan area and for other urban areas appear in Tables 2, 3, and 4 in Appendix 1.

### 3.5 Selection of Households from a Ward by CSSEQP for Individual Questionnaire

As already stated earlier, the Census documents relating to the sampled wards and sub-wards should be extracted from the Census records. In case the households enumerated in the sampled wards have not been serially numbered and ordered sequentially, a sample reference number should be assigned to each household. The last number assigned should be equal to the total number households in the ward or sub-ward  $M_{hi}$  enumerated in the Census.

Circular Systematic Sample with Equal Probabilities of selection, **CSSEQP** is recommended in the selection of households from a selected ward is a simpler version of **CSSPPS**. To select a sample of 15% of the households enumerated in the Census compute the number  $M_{hi} \times 15 / 100$  to the nearest integer.

The sampling interval  $I = M_{hi} \times 100 / M_{hi} \times 15 = 6.67$  say 7 when rounded off to the nearest integer. A random number  $R_{hi}$  in the interval 0 to  $M_{hi}$  should be taken from a book/sheet of random numbers as the first selector number. The remaining selector numbers should be calculated one after the other by adding  $I = 7$  to the previous number. If the sum exceeded  $M_{hi}$  the remainder after subtracting  $M_{hi}$  from the sum should be taken as the selector number. These selector numbers are the serial numbers of the selected households. The census schedules bearing the selector numbers are the sampled households. This process should be repeated for all the sampled wards. Selection of households from the sampled villages should be done by staff of the Population Section/ PES supervisors.

### 3.6 Estimation Method

The dual system of estimation is based on the capture and re-capture (Chandarasekaran –Deming) model where a sample is taken from a population and tagged and released and counted in a re-capture.

Assume that  $N$  = the true population size, the value to be estimated  
 $N_c$  = the population enumerated at the census of population; and  
 $N_p$  = the total estimated population derived from the PES  
 $M$  = the total number of persons captured in the PES and are matched against the census;

Then, the total population  $N$  is given by

$$N = \frac{N_p \times N_c}{M}$$

That is the total population is estimated by the number captured in the census times the inverse of the coverage rate of the census as measured by the PES. This model assumes that matching of persons and units is undertaken perfectly between the census and the survey; that there are no multiple observations of individuals in either source and that the two sources are independent.

This estimation can be improved by adjusting the counts for errors that arise during enumeration. These include

- $G$  = the number of persons incorrectly located geographically in the census.
- $E$  = the number of persons incorrectly enumerated in the census; and
- $D$  = the number of duplicate enumerations in the census
- $I$  = the number of persons who are enumerated in the census, but having insufficient information for matching

The estimate of the total population size  $N$  after adjustment for these errors in enumeration is given by

$$N = \frac{N_p (N_c - G - E - D - I)}{M}$$

where

- $N_p$  = the weighted total estimate of persons in the P sample ( Form 1)
- $N_c$  = the total number of persons enumerated in the census
- $G$  = the weighted estimate of persons miss assigned geographically in the census from the E sample (Individual Questionnaire-Form 2)
- $E$  = the weighted total estimate of persons erroneously enumerated in the census from the E sample (Individual Questionnaire -Form 2 (e.g. infants born after census day)
- $D$  = the weighted total estimate from the E sample (Form 2) of duplicate enumerations in the census.
- $I$  = the number of persons who were enumerated in the census, but having insufficient information for matching from the E-sample (Form 2)

The coverage rate,  $O_n$  can be estimated as the total census count as a proportion of the total population size ie. the observed census total divided by the dual system estimate of the true population size.

$$O_n = \frac{N_c}{N} = \frac{N_c}{N_p(N_c - G - E - D - I)/M} = \frac{M}{N_p} \times \frac{N_c}{(N_c - G - E - D - I)}$$

$$= \frac{M / N_p}{(1 - (G + E + D + I) / N_c)}$$

$$= \frac{1 - Gu}{1 - Gu}$$

$$= \frac{M / N_p}{(1 - (G + E + D + I) / N_c)}$$

where  $M / N_p$  is the estimate of the coverage rate and the denominator is the correct enumeration rate

$$= \frac{1 - Gu}{1 - Gu}$$

where  $Gu$  = the gross undercount rate uncorrected for error in the census and  
 $Go$  = the gross undercount rate including census errors.

Accordingly, the under-count rate can be expressed as

$$U_n = 1 - \frac{N_c}{N} = \left( 1 - \frac{N_c}{N_p(N_c - G - E - D - I)/M} \right)$$

$$= 1 - \frac{M}{N_p(1 - (G + E + D + I) / N_c)} = 1 - \frac{M(1+F)}{N_p}$$

Using Taylors expansion  
where  $F = (G + E + D + I) / N_c$

$$= \frac{(N_p - M) - M * F}{N_p} = \frac{(N_p - M)}{N_p} - \frac{M}{N_p} \times F$$

That is the out of census population divided by the PES estimate less the proportion of erroneous population multiplied by the coverage rate. The last term was not computed as the CBS could not process the erroneous enumeration sample data in Form 2.

The computation of undercount rate was carried out within the four strata. The inflation factors were applied to the census and survey counts determined for each sample

ward or sub-ward in the cluster sample where total enumeration was carried out within the sampled ward/sub-ward.

## **4.0 Field Operations**

### **4.1 Field Test**

The draft questionnaires and instructions to enumerators were discussed internally and amended. The pre-tests were conducted in Chapagau VDC in the Lalitpur district and in the Kathmandu Municipality Ward 21 and Ward 14 on 25th and 26th of September 2001 by a team of seven enumerators drawn from the staff of the Population Section of CBS. The senior officials of the Population Section had participated and supervised the conduct of the pre-tests. The staff functioned in two groups in conducting the tests and there was good cooperation from the public. The pre-tests involved the identification of ward level boundaries in terms of instructions, and administration of the two questionnaires to selected households. The tests were useful in identifying several deficiencies in the survey questionnaires and instructions. The draft questionnaires and instructions were modified on the basis of these tests, these modifications related to the wording of questions, space provided for responses, skip instructions and modifications and clarifications of concepts and instructions in the enumerators manual. The draft questionnaires were finalized on 2<sup>nd</sup> October and the Listing Form and the Individual questionnaires were printed on 5<sup>th</sup> October 01.

### **4.2 Training of Field Staff**

All together 60 staff from the District Statistical Offices and Headquarters had been selected for training and deployment on the survey. In order to ensure control over quality to the extent feasible in the local context, it was decided to conduct training centrally although it would raise costs and was logistically more difficult. The centralized training of field staff was supported by UNFPA. The field staff were trained for three days commencing 8<sup>th</sup> and ending on 11<sup>th</sup> October 2002 with a practical training exercise on field enumeration. The training was conducted at the Administrative Staff College, Jawalakhel by CBS senior staff at headquarters who functioned as resource persons. The training was organized as lecture sessions in the morning where the entire group attended the sessions. The group was divided into 4 sub-groups and the afternoon sessions were conducted as group discussion classes. The two questionnaires and the instructions to fill them and selection of sample households for canvassing the individual questionnaire were explained and discussed during these sessions. The class-room training was followed by a field visit to Dakshinkali (Pharping) on 11<sup>th</sup> October 01 where the trainees conducted interviews under the supervision of the senior staff of the Population Section. The selection of staff who had previously worked on census and household survey field listing and data collection operations made it possible to reduce the training duration and also to focus training on conceptual and procedural aspects, specifically concerned with PES enumeration. Detailed instructions to fill the household listing form and the individual form were prepared which could be used as training and reference material during training of enumerators and supervisors and subsequently as reference material during the field survey. See Annex...

### **4.3 Enumeration**

As far as the PES enumeration procedures were concerned they were basically similar to those adopted in the census. The census material prepared during the enumeration phase of the census in respect of the wards or sub-wards that were sampled for the PES had to be identified, located and extracted from census records. These materials included the maps or sketch maps that detailed significant land marks and boundary descriptions, household listing forms, and household enumeration forms etc. The fieldwork on the PES consisted of three distinct phases.

- i. identification of sampled ward boundaries and listing and enumeration of all households using PES Form 1;
- ii. re-enumeration of the usual resident members of sample of households drawn from census records in the sampled ward using PES Form 2;
- iii. matching of information collected in the PES with those contained in the census household schedule and individual schedules in office;
- iv. field verification of information of households and individuals where PES information did not match with census forms.

Of these activities the first two were conducted simultaneously as fieldwork on the PES. In order to reduce the workload that arise in field verification, which is time consuming and costly, the collection of accurate and complete information in PES field data collection stage was essential. This need was emphasized in the training sessions.

The fieldwork of listing of households using Form 1 and canvassing of Form 2 commenced on 15 October 2001 and and completed at the end of 18 November 2001. The deployment of two trained interviewers to work together was accepted as desirable to expedite the enumeration work and also to reduce errors through a single interviewer over looking some important information. The use of two enumerators in the PES had distributed the workload, prevented double counting of households, reduced the time taken for the enumeration and also improved the quality of data collected. The District Statistical Officers (DSO) of CBS supervised the fieldwork of enumerators. DSO's had checked the enumeration work by carrying out surprise inspections, so that any work of doubtful quality can be checked and adjusted in the field itself. They had also provided logistical support to the enumerators.

### **4.4 Household Interviews**

As in the census, the fieldwork on the PES also consisted of two distinct phases. In the first phase the sample enumeration area/ ward had to be correctly identified in terms of the land marks and boundaries shown on the map or sketch map prepared at the blocking out stage of census enumeration areas or with the help of village level officials. The trained enumerators were required to meet the local level village officials and village elders and clearly identify the boundaries of the sampled ward. A careful identification of the ward boundaries especially in the rural areas was accepted to be a difficult task and the field staff were advised to check with knowledgeable persons and householders themselves residing in the vicinity of the ward boundaries to accurately demarcate the boundaries of the ward.

In spite of these precautions, there were difficulties in correctly identifying the boundaries. In a number of cases, the boundaries of the wards and sub-wards identified by the PES enumerators have not matched with those used by the census enumerators resulting in problems of reconciling records. This problem was a major issue in Kathmandu and other urban areas that impaired the coverage and quality of the data collected. The data collected in Kathmandu Ward 34 Sub-ward 20 could not be matched as the boundaries did not correspond with those used in the census enumeration. Less serious issues were encountered in other areas too. The enumerators and supervisors should have noticed this problem of lack of correspondence of boundaries between the census and the PES wards as they had to identify a number of sampled households drawn from the households enumerated in the census whose identification particulars were given to them at the time they were deployed on enumeration. The fact that the boundaries of the sampled wards and sub-wards had not been correctly identified is also evident from the presence of a large number of households enumerated in the census but not enumerated in the PES. These problems have of course arisen primarily as a result of a lack of maps and sketch maps to guide the enumerators and these issues will continue to trouble the CBS until census cartography is improved and strengthened.

The listing operation commenced after the identification of ward boundaries where each household in the enumeration area had to be visited to collect the data in PES Form 1. At this stage basic particulars relating to the housing unit and household within the sample ward had to be collected to uniquely identify each household. Information was canvassed by visits to households and face-to-face interviews with one or more senior members of the household. The household size at the time of the census and on the day of the interview broken down by sex and changes in household headship and composition after the census were also collected. The personal details canvassed from the household head and the eldest household member included the person's name, surname, sex, and age. Usual procedures adopted in the house listing operations were used to ensure methodical listing of all households and canvassing of particulars from the household head and other senior members of the household.

#### **4.5 Individual Questionnaire -Form 2**

The individual information questionnaire (Form 2) was used to collect information from the sampled households in the sampled ward. The questionnaire elicited information about all persons who resided in the household on census day. The questionnaire verified whether the respondents were in fact counted in their households at the census. The information on name, surname, sex, age, and usual place of residence and status of enumeration at the census was collected for each person in the sampled household. In addition, information on changes in household composition due to migration and deaths was also canvassed.

The difficulties of contacting the household were the most common reason for not completing the questionnaires as well as for incomplete entries recorded in Form I and Form II. There were no reports of refusals to participate in interviews. However, security concerns had started to become issues and there had been reluctance to undertake field work in some areas such as Accham district.

#### **4.6 Follow up Interviews**

These interviews were mainly concerned with the issue of a large number of unmatched households and persons in Form 1 that could not be resolved in office. In the urban areas it was found to be mainly the result of non-overlapping boundaries that caused differences in the enumeration area surveyed in the census and the survey. This was the main issue in Kathmandu, to a lesser extent in other urban areas and it also affected the rural areas too. The re-interviews of unmatched households could be undertaken in only about 50% of the sampled wards. Hereto field verification had to be restricted to check omissions, differences in household sizes, geographically miss-located households but investigation of omitted persons using Form 1A, Form 1B and Form 2B was not followed through. The field checks had made it possible to resolve a number of issues relating to unmatched records and to classify them.

The difficulties encountered in resolving issues concerning matching PES Form 1 with the census records made it necessary to give priority to finalize the processing and matching of Form 1. The work on this item continued up to the 3<sup>rd</sup> week of February. As a result the processing of Form 2 had been delayed and the conduct of re-interviews not meaningful. Since the people are mobile, and more so under disturbed conditions, it is important to conduct the follow up interviews without delay. The disturbed situation in the districts referred to earlier was also a major reason for delaying the immediate follow up on cases that required re-interviews. Staffing was another issue at a time when the CBS was fully involved in population Census processing and on going fieldwork on the National Census of Agriculture.

#### **5.0 Matching of PES and Census Records**

The post enumeration survey methodology for evaluating census results depend on the matching of PES records with those of the census and determining records that match, that do not match and those that cannot be determined with the available information. The process of matching was carried out by comparing information given by respondents in the PES with information given in the corresponding census forms. Objective of matching is to make a determination whether the respondent in the PES was counted at the census at each address.

##### **5.1 Matching Rules**

Detailed rules for matching were prepared using a set of characteristics used in canvassing the PES and the census for the training and guidance of staff entrusted with matching and searching operations. The matching rules specified the characteristics: name, age, sex, etc. by which persons enumerated at the PES and the census should be compared to establish a match. The tolerance ranges within which records must agree was also defined and recorded. The matching rules were prepared to conduct the matching operation in 2 or 3 stages as required, assigning appropriate staff for the different stages. In the first stage, cases which were apparent and clear where a definite match or a definite non-match were obvious were to be disposed of by trained junior staff and less clear and doubtful cases were to be tackled by supervisory level staff responsible for matching operations in subsequent stages. ( See Annex .....)

In devising the rules for matching, erroneous non-matches i.e. cases accepted as non-matches although a particular person corresponds to the case enumerated in the census should be minimized, and this was considered in framing the rules. However, when erroneous non-matches are kept low the erroneous matches where a person who was not enumerated at the census is accepted as having matched with a record of a person enumerated at the PES will be increased. The objective of the matching rules is to reduce both erroneous non-matches and erroneous matches to the extent feasible.

## **5.2 Matching of PES Form 1 and Census Records**

In all 6 staff including 2 computer operators were trained in matching the PES records against the census records and were deployed as record matching staff who worked under the guidance of the UNFPA National Consultant. The first stage of matching involved the location of census questionnaires corresponding to the PES, ward or sub-ward. The identification and location of the relevant census documents was time consuming in some wards and sub-wards where the boundaries had not been demarcated clearly and the area listed in the PES did not match exactly with the area enumerated in the census.

The matching staff used a computer print out of the re-ordered list of household heads sorted by surname in alphabetical order to match the PES records against the census records. Matching operation was undertaken by comparing whether the name of the head or another member of the household included in the census form was included in the re-ordered list PES heads of households. The personal details relating to name, age and sex were also available for comparison. These variables did not have to be exactly the same for a match to be made as often particularly for age responses are likely to differ. Where this was inconclusive the questionnaire was compared to see whether a match could be based on information on eldest member of the household name, age, sex etc. For doubtful cases other details collected in the PES and the census were also available. Usually when a corresponding census questionnaire was identified all members of the census household were classified as enumerated or missed. In the event of a match the household serial number of the census schedule was entered in PES Form 1 adjacent to the corresponding household number in Form 1. The individual form (Form 2) was not processed, the entries in the PES individual Form 2 could not be matched with the entries in the census individual schedule for the reasons stated earlier.

## **5.3 Matching Issues**

Experience in conducting the PES 1991 was also useful in avoiding some of the issues involved in matching of records. Specially because of the rather long gap of 3-4 months between the census and PES the possibility that the head of the household may have changed was anticipated and the questionnaire has elicited information on the name of the head of household at the time of the census. Similarly changes in the household composition during the intervening period between the census and the survey were also canvassed in PES Form I. There were several problems that were observed in matching:

- i. cases where the name of household head and eldest member of household were spelt differently in the PES and the census forms;
- ii. household head had changed between census and the PES;

- iii. nuclear households had consolidated or vice versa or consolidated households had been divided and reported as such at the census or PES;
- iv. the occurrence of more than one household in a ward with the same name of head of household and caste.
- v. consolidation of nuclear households with those of their parents for a short time during festivals such as Desain and Tihar when PES field work was ongoing.

These and other issues had resulted in a relatively large number of non-matched households in the first tier of matching operations. When the rules were relaxed and the matching operations were conducted it was possible to resolve some of the problems of matching and achieve matched status. Some of the issues had to be resolved through field visits and re-interviews. The use of the printout with names in alphabetical order facilitated the matching of records especially when the sampled ward had a long list of households to match. The use of the computerized list also reduced the incorrect matches as all names with minor spelling differences were available for consideration and also increased the matched records as it enabled other criteria including household size, sex and age also to be taken into account in arriving at a decision.

The matched Form 1 and the census records were reviewed case by case and each household was classified as matched, missed or unresolved. In about 50% of the sampled wards, after initially reviewing the doubtful cases they were referred for re-interviews before the list of records was prepared. Finally, UNFPA National Consultant reviewed the records and the list of matched persons, and unmatched cases broken down by out of census records, and out of PES records were compiled. The aggregate data for the ward was compiled which was used in estimation.

## **6.0 Data Processing**

PES data was processed at CBS Headquarters. PES questionnaires were checked and taken over by the staff of the Population Section. The preliminary checks for completeness of PES questionnaires were carried out at the stage of taking over of documents. In order to facilitate matching operations a decision was taken to key in the information in the Listing Sheet Form 1, considering the difficulties encountered in matching records manually in PES 1991. IMPS software developed and supported by the US Bureau of the Census was used in data entry operations. To use the data files for matching purposes it was necessary to input the names and surnames of respondents. At this stage, some initial checking and editing of entries in the questionnaires were also undertaken while translating the names into English. The information in Form 1 which are needed for matching and estimation purposes including household serial number, the number of male and female household members reported to have been enumerated in the census, the number of males and females listed in the PES, the name and surname of household head and the person's sex and age and the same information for the eldest member of household were keyed in. Thereafter the data file for the sampled ward or sub-ward was checked for completeness and the data file was sorted alphabetically by name and then by surname of the head of household. The matching staff used a print out of the re-ordered list of household heads in alphabetical order to match the PES records against the census records.

The need to enter the name and surname of heads and eldest members of households increased the volume of data entry work load of the two operators assigned to the PES and the keying in of information in the listing sheet Form 1 was continued until the 3<sup>rd</sup> week of February 02. The delays in data entry and matching of Form 1 blocked other activities, the individual questionnaire –Form 2 could not be processed and matched. The data in Form 2 had not been checked and keyed in. The field follow-up of unmatched and doubtful cases in Form 1 could be undertaken only in about 50% of the sampled wards. The security situation had also deteriorated by this time to undertake follow up re-interviews in a number of sampled areas. The delays in processing and matching also did not permit the creation of the census records file, the matched records file and the file containing out of census records for processing and preparation of PES estimates. The aggregates of matched records that were prepared were used with excel software to set up spread-sheet programs to produce computations and tables.

### **6.1 CENVAR Software**

CENVAR is the variance computation module of the IMPS software, and it is designed to produce the estimated values of the parameter, the standard error, the coefficient of variation, the confidence interval and the design effect. The manual on CENVAR contains an illustrative case study on PES that had adopted the dual system estimation procedure. CBS should arrange to obtain this software, which is distributed free to developing countries by the US Bureau of the Census. The processing of variance computations can be messy and time consuming and the acquisition of this software and training CBS staff in its application are therefore recommended.

## **7. Estimation**

The results from matching were used to calculate the number of households and persons who were counted in the PES, the number that matched with the census, the total number enumerated in the census less any adjustments effected; the number enumerated in the PES but omitted in the census; and the number enumerated in the census but omitted in the PES. These aggregates for each sample point were used in the estimation procedure with the corresponding weighting factors for the sampled clusters to derive the estimates of the variables for the domain. The Excel spread sheet program was used in the computation of coverage and undercount rates and the precision of the estimates. The computation procedures, the tables generated and results in Excel spread sheets are copied into the computers in the CBS Population Section.

### **7.1 Sample Data**

The PES 2001 had surveyed 8,188 households with 42,692 household members. All the sampled wards were surveyed, but there were problems of non-overlapping boundaries discussed earlier, and one of the sampled wards had to be dropped and the inflation factors had to be adjusted. Similar adjustments to the weights were necessary where only partial coverage was achieved in two urban sub-wards due to boundary definition issues and in the Accham district mainly due to security concerns.

The number of households listed in PES sampled wards, the number of persons in the listed households, the number of persons survey sample that matched with the census and the number out of the census count are shown in Table 2.

Table 2  
The number of Households and Persons survey in

Region/Domain	Wards	Number of Households	Number of Persons		
			Listed in PES	Matched	Out of Census
Terai	16	3247	18249	17286	963
Hill&Mountain	30	3228	16143	14916	1227
Kathmandu	4	720	2819	2319	503
OtherUrban	4	993	5481	4609	872
Nepal	54	8,188	42,692	39,130	3565

The breakdown of this information for the sampled wards in each stratum is given in Table 2 in Annex 5. The inflation factors were used to blowup the sample data and estimates were aggregated at stratum level. These sample data with the weighting factors were used to derive the estimates for the different strata. The stratum level estimates were aggregated to derive regional and national level estimates.

The PES sampling design was able to capture the significant features of the population structure and distribution, despite the fact that sample selected was rather small in relation to the samples usually selected for post enumeration surveys. This is seen from the stratum level population estimates from the PES which are fairly close to the enumerated population in the census for the regional break downs.

Table 3  
PES Estimates and Census Counts for the Strata

Sector	Census 2001 Enumerated Population	PES Sample Estimate	Difference	
			Number	%
Nepal	23,214,681	22,841,118	373,743	1.6
Rural Sector	19,911,181	19,493,608	417,573	2.1
Tarai	9,744,810	9,999,055	-254,245	-2.6
Hill and Mountain	10,166,371	9,494,553	671,818	6.6
Urban Sector	3,303,500	3,492,607	-189,107	-5.7
Kathmandu	696,852	740,184	-43,332	-6.2
Other Urban	2,606,648	2,752,423	-145,775	-5.6

## 8.0 Results

The PES estimated the gross undercount rate for Nepal at the Census of Population 2001 as 0.0644 which means that 6.44% of the total population of the country was omitted in the census count. This rate is substantially lower than the omission rate estimated for the Census 1991 which was as high as 12.38%.

The reported omission rate is the gross undercount rate and it does not take into account the over-count of the population due to erroneous enumeration such as double counting etc. Therefore the survey estimate is an upper bound of the under-enumeration. The over-count rate could not be computed as the CBS had found it difficult to match the individual population sample and conduct the re-interviews due to staffing and security problems. The PES undertaken in 1991 had shown that the over-count in the Census 1991 was 1.77% when the undercount rate was estimated as high as 12.38% leading to a net omission rate of 11.05%.

The above omission rate is the gross omission rate arising from the failure to enumerate eligible persons for whatever reasons. Thus, the factual position is that approximately 1 out of 16 persons have been omitted from the census. The fact that other persons may have been double counted or erroneously enumerated does detract from the position that the omitted persons had not been counted in the census.

The sampling variance of the gross undercount rate was computed and it amounted to  
with a standard error of the estimate of  
and its coefficient of variation was

$\text{var}(r) = 4.83\text{E-}5$
$\text{se}(r) = 0.0069$
$\text{cv}(r) = 10.79$

These values of the standard error and coefficient of variation of the omission rate are satisfactory in relation to the relatively small sample of 8,000 households sampled in the survey using a cluster sampling design to reduce both field work and survey costs.

According to the provisional estimate of Census 2001 the total enumerated population of Nepal is 23,214,681. The gross undercount rate represents an undercount of about 1,495,000 people. In terms of these results the total population of Nepal would have been 24,710,000 had the total population been completely enumerated. Further, according the precision of the survey estimates the total population of Nepal would have been between 24,387,000 to 25,033,000 people 19 out of 20 times if all the omitted persons had been enumerated.

Table 4  
National and Regional Undercount Rates and the Precision of the Estimates

Sector	Census 2001 Enumerated Population	Coverage Rate	Omission Rate(r)	Standard Error Se(r)	Coefficient of variation cv(r)
Nepal	23,214,681	0.9355	0.0644	0.0069	0.108
Rural Sector	19,911,181	0.946	0.0539	0.0046	0.0853
Tarai	9,744,810	0.952	0.0474	0.0059	0.126
Hill and Mountain	10,166,371	0.939	0.061	0.0071	0.117
<b>Urban Sector</b>	3,303,500	0.877	0.1222	0.034	0.270
Kathmandu	696,852	0.890	0.109	0.067	0.611
Other Urban	2,606,648	0.874	0.125	0.039	0.311

### 8.1 Geographic Coverage

The census coverage had varied across the regions and that should be expected in Nepal. The undercount rate was lowest in the rural sector segment of the Terai ecological zone which amounted to 4.7%. The standard error of the estimate was also low which produced a percentage relative standard error (PRSE) of 12.6. The sample size was not large and the rural part of the Hill and Mountain ecological zones were treated as one stratum. The combined Hill and Mountain had a higher undercount rate of 6.1%. More difficult terrain and the difficulties of mounting efficient census operations comparable to that undertaken elsewhere would have raised the undercount rate in these regions. The use of 30 sample wards had slightly lowered the PRSE to 11.7 when compared with the rate of 12.6 for Terai which had a lower undercount rate.

The combined rural sector comprising Terai, Hill and Mountain ecological zones had a omission rate of 5.4% and a PRSE of 8.5 as the two strata taken together had 46 sample wards which had contributed to lower the standard error of the omission rate. The omission rates are higher in urban areas which amounted to 10.9% in Kathmandu rising to 12.5% in other urban areas. The omission rate is higher in urban areas as people are more mobile in urban areas than in the rural sector, and consequently they more likely to be omitted and miscounted in places other than their own usual place residence. The lack of census maps and sketch maps had also contributed to raise the undercount rates with the persons residing close to the sub-ward boundaries which are not properly identified being excluded or mislocated and lost in census and PES enumerations.

The overall undercount rate for Nepal is determined primarily by the rural sector rates in Terai and Hill and Mountain regions, as the share of the urban population amounts to only 14% of the total population.

In this dual system estimation approach, the omission of households in the PES has had the effect of raising the undercount rate. The PES household listing operation had omitted households enumerated in the census, and these households had not been followed through in the field follow-up through re-interviews, to verify whether they were geographically mislocated or otherwise erroneously enumerated households. Such follow up is of course is time consuming and costly. It is evident from the formula that the coverage rate

$$O_n = (N_c / N_c * N_p / M) = M / N_p$$
would have increased had the PES enumeration was more complete, then the undercount rate would in fact been lower than what is reported through these estimates.

## **9.0 Conclusions and Recommendations**

### **PES 2001**

Mainly as a result of the responsibilities connected with the population census and the on going National Agricultural Census 2002 and some routine statistical activities it had not been possible for the CBS with staff constraints to give sufficient attention and planning to the PES. The work on the PES had really commenced after census enumeration. As a result many stages in the planning, designing and operationalizing had to be telescoped so that the PES enumeration could be undertaken within the shortest possible date after the census enumeration. The proper planning and designing of a PES needs at least 1 to 1 1/2 years and therefore the work on future PES should commence at least 1 to 1 1/2 years before a census enumeration. This would allow various survey procedures to be tested and reviewed; the methodology and documentation modified as required; and systems and programs prepared for survey processing and estimation. The very important requirement of conducting the PES immediately after the conclusion of the census had not been observed hitherto. Unless the fieldwork on the PES is undertaken immediately after the census enumeration the quality of the data are bound to suffer because of mobility and memory lapses.

9.1 The PES 2001 was an advance over the PES conducted in 1991, most importantly in undertaking field work in all the sample enumeration areas which was not achieved in both previous PES surveys conducted in 1981 and 1991. In addition fieldwork was undertaken by centrally trained staff and matching operations were conducted in a more organized manner. On the other hand a number of factors had affected the outputs from the survey and it is important to note these limitations in order to avoid these deficiencies and to improve procedures in the future. The sample size had to be restricted in view of both funding and manpower constraints. The sample size chosen was barely adequate to realize statistically reliable estimates at aggregate and sectoral levels. Yet the CBS found it hard pressed to provide staff inputs to meet the planned outputs from he survey. In both PES 1991 and PES 2001 to reduce the work load household was treated as the unit of enumeration and information for each person was not canvassed but instead a listing sheet was canvassed from the households. The re-interviews could not be conducted as planned and the individual questionnaire – Form 2 could not be processed and as a result the over-enumeration rates could not be estimated.

9.2 The PES should be planned as an activity of the census programme coordinating it with the census so that data collection on PES can be undertaken within a month from the date of census enumeration. To achieve this objective two conditions will have to be realized, firstly the planning and designing of the PES should be finalized possibly immediately after the pilot tests for the census are conducted and census enumeration documentation are finalized. Further, the PES questionnaires and instructions manuals should be prepared and field tested and finalized. It is also essential to ensure that the systems and programmes for PES processing be prepared test checked and finalized. Thus, a few weeks prior to the census enumeration PES plans and procedures should have been completed and a programme of training of PES enumerators and supervisors finalized including selection and deployment of staff.

9.3 Staffing had been a major constraint that had affected the quality of the PES that had been hitherto undertaken. Thus, recruitment, selection and deployment of staff for the PES should be organized as a separate project activity. Because of census operations, if the release of staff is an issue, then other arrangements will have to be made drawing only core staff from CBS and other cadre requirements from elsewhere.

9.4 The second issue that had seriously affected the PES has been funding constraints. Any census evaluation programme will require the conduct of a post enumeration survey of adequate scale and dimensions to support the type of checks on sectoral and regional breakdowns and demographic characteristics that should be investigated. The sample size needed for such an investigation can be determined and the cost estimates for the survey prepared. Once the cost estimates are prepared it would be possible to seek funding from bilateral or international agencies to support the financing of the PES if there are constraints in securing funding from government budgetary sources. If the funding is required under technical assistance programme, then the proposal for financing should be prepared well in advance to ensure that funding will be received in time to conduct these operations. The PES should be planned and designed well ahead of the census enumeration so that the data collection on PES can be undertaken within a few days after census enumeration.

9.5 The undertaking of a large PES immediately after the conclusion of a census is a difficult task and this problem must be recognized and the strategy for implementing the PES will have to be carefully designed and coordinated with the census. In order to make a meaningful evaluation of the census coverage the PES sample should be sufficiently large to provide statistically reliable estimates for the urban and rural sectors and other principal geographic breakdowns. If the content errors are also to be estimated then the sample size has to be further strengthened. The matching of records and field verification of doubtful cases is both time consuming and require staff support. Thus the resources required including staff cadre for different stages of the PES the data processing plans and procedures and cost of other inputs will have to be carefully assessed. In addition there are several other matters that should receive attention:

- development of good area frames with well defined and mutually exclusive enumeration areas;
- appropriate sampling design which would facilitate objective

- generalization of PES results to relevant domains; and
- formulation and adoption of realistic matching rules/criteria.
- Allocate adequate funds for the PES within the framework of the census.

## **9.6 Census Cartography**

The lack of cartographic materials for census mapping is undoubtedly the main deficiency that had affected the completeness and quality of census and survey data. In the absence of cartographic materials it would be difficult to use the sampling frames based on the Census 2001 for inter-censal household and establishment based surveys. The main issue that the PES encountered in enumeration was the lack maps and sketch maps for the census enumeration areas. The CBS should establish a Census Cartographic Unit with basic equipment and facilities as soon as possible securing technical assistance funds. A cadre should be established and 2-3 statisticians should be trained in census mapping. A few officers who have the aptitude for cartographic work should be identified and trained in a national statistical organization such as the US Bureau of the Census or the Australian Statistics Bureau. The field staff deployed on surveys should be detailed to prepare a sketch map of the sampled enumeration area in all sample surveys to reduce under-enumeration and miscounting.

## **9.7 Sample Survey Unit**

The CBS should strengthen Sample Survey Section to support the technical work on household and establishment based surveys. CBS should arrange to train a few statisticians in planning and designing sample surveys and in survey sampling. Two officers who have strong mathematical/statistical backgrounds should be identified and provided with post-graduate training in survey design and survey sampling. CBS could request for this training under technical assistance programmes. At present the paucity of trained staff prevents the CBS from benefiting from the statistical surveys currently being undertaken and from the work presently being done utilizing the services of expatriate staff recruited for different project tasks.

## **9.8 Sampling Design**

The sampling designs adopted in PES have varied in different countries. There are some advantages in using existing sample designs adopted by the country as it could reduce enumeration costs of field data collection since the interviewers are already familiar with the geographic areas, and cartographic materials, etc. are also available for those areas. The recently conducted PES (1996) in New Zealand had used a two stage stratified cluster sample from the sampling frame used for the labour force survey. The total sample had amounted to approximately 0.8% of all private dwellings. In Nepal where cartographic materials are a major issue, a sampling design that had been used within a year or two prior to the planned PES and produced good estimates may be considered for use in a PES that will be planned in the future.

## **9.9 Adjustment of Census Estimates**

After each census CBS has produced population estimates and projections for Nepal and sub-national areas. In the construction of these estimates the enumerated population at the census had been taken as the base population. It appears that the

undercount at the census, had not been taken into account in adjusting the base year population. The inter-censal population estimates and projections will be more complete and reliable if the population adjusted for the undercount in the census is taken as the base population. The information relating to the PES sample data are available in the computer files that were generated for the preparation of estimates included in this report.

### **10.0 Acknowledgements**

I would like to express my sincere thanks to Mr. Keshav Raj Sharma, Director General, CBS, Dr. S. M. N Suphani, Deputy Director General and the members of their staff for their co-operation and assistance. I would also like to thank Mr. Radha Krishna G. C. Deputy Director, Population Census, Mr. B. D. S. Dangol, UNFPA National Consultant, CBS and Mr. H. N. Regmi, UNFPA National Data Processing Consultant for their advice and useful comments during my mission and in the preparation of this report.

Central Bureau of Statistics,  
Kathmandu  
February 2002

### **References**

1. US Department of Commerce, Bureau of the Census, Evaluating Censuses of Population and Housing, Washington, 1985.
2. Natarajan, K.S., Report on the Coverage of 1991 Census of Nepal, CBS, Katmandu, 1993.
3. Dangol, B. D. S, PES 2001, CBS, Kathmandu, 2001.
4. Korale, R. B. M. Post Enumeration Survey 2001- Draft Preliminary Report, CBS, Katmandu, Sep. 2001.