# SAMPLE DESIGN FOR THE 2001 CENSUS OF HOUSING AND POPULATION, NEPAL

Kwok Kwan Kit
*Sampling Consultant*

# 1 Introduction

The scope of population censuses has increasingly been enlarged to include a wide range of topics to satisfy the increasing demand for data. A new interest in a range of social data has also emerged, principally from a greater awareness of several social issues, such as gender issues, the situation of children and youths, the elderly and people with disability. However, a 100 percent enumeration of all included topics places too heavy a burden on the census. This is likely to result in poor quality data and delay in data release because of the sheer volume of data to be collected and processed. A cost-effective approach that can be adopted to accomplish the objective of expanding the scope of the census without overburdening the census is to obtain some of the required information from a 100 percent enumeration and the rest from a sample.

The Central Bureau of Statistics (CBS), Nepal is currently planning the next population census to be held in the year 2001 and will, for the first time, be adopting the use of sampling in the enumeration phase of the census in order that the 2001 census will be able to provide the comprehensive range of information required by users.

The main objective of this report is to provide a description of the sample design for the 2001 Census of Housing and Population, Nepal. In addition, the procedures for: (a) selecting the sample, (b) computing estimates from the sample and (c) computing standard errors are given in detail to facilitate implementation.

The report is in five sections. An introduction is given in Section 1. Section 2 provides an account of several sampling methods that can be used for census enumeration. Section 3 gives a description of the main features of the proposed sample design, while Section 4 documents the steps taken to arrive at a suitable sample size. The estimation procedures, including the computation of ratio estimates and the use of the random group method for the estimation of sampling errors, are given in Section 5.

# 2 Alternative Sampling Methods for Census Enumeration

Once the decision to use sampling in the census enumeration is taken, the different sampling methods that can be used need to be examined for their operational feasibility. The procedures to be adopted need to be included as part of the overall census enumeration plan, including the collection of the complete counts.

Basically two broad approaches are available:
(a) Sampling of whole enumeration areas (EA) and
(b) Sampling within every EA.

## 2.1 Sampling whole EAs

In this approach, whole EAs are selected and all households and persons in the selected EAs are enumerated using both the basic and the supplementary questionnaires. Households and persons in the non-selected EAs will only answer the basic questionnaire.

In past censuses in Nepal, the ward, which is the smallest administrative sub-division, has been used as the census EA though in the municipalities areas or urban areas these had to be subdivided into several EAs because of their large size. It is estimated that there are about 36,023 wards in the country and assuming that, on average, the Municipality wards are subdivided into 6 EAs gives a total of some 40,000 EAs. Since the sampling unit used in this approach is the EA, an areal unit, the sample can be called an area sample. Also since the EA is a cluster of households and persons, the sample can also be called a one-stage cluster sample.

There are certain advantages using this area approach, the chief is probably that the sample selection of EAs can be done by the CBS, thus avoiding selection biases when sample selection has to be carried out in the field.

Other advantages include:

a) A limited number of enumerators need be trained to ask the supplementary questions since the supplementary questionnaire has only to be canvassed in the sampled EAs.
b) Filling in the census forms is facilitated since the enumerators in the selected EAs use both the basic and supplementary questionnaires for all households in their EAs.
c) Editing, coding and tabulation operations can also be simplified.

However, this approach is statistically less efficient compared to sampling within every EA. The average number of EAs per district, is rather small in Nepal. On average, this is 533, though some districts contain considerably more EAs than the average and some much less. Sampling EAs and enumeration of all households and persons would result

in large clustering effects, indicated by the Design Effect (deff) which is in this case, the ratio of the sampling variance of the one stage cluster design over that of sampling within all EAs with the same number of households or persons used in the first design. For most socio-economic variables, deff is greater than 1.0 indicating that the cluster design has a larger sampling variance and thus less efficient. The design effect can also be shown to be of the following form which is informative:

$$deff = 1 + roh(\overline{B} - 1) \quad \text{..... (1)}$$

In (1) we see the size of deff is influenced by the rate of homogeneity (roh) a measure of cluster homogeneity, and $\overline{B}$, the average cluster size. For most socio-economic variables, roh has been observed to take small and positive values, thus contributing to making deff greater than 1. For estimates based on the total population, $\overline{B}$ is the average population size of EAs in the district. An averaged - sized district could have 300,000 persons by 2001 and this will result in a $\overline{B}$ of 563 persons per EA. Even for a fairly small value for roh of 0.04, we get a large design effect of 25. Estimates from the sample based on smaller subclasses will have smaller design effects. For household characteristics, $\overline{B}$ is the average number of households per EA. For example, the average number of households per EA in an average sized district could be about 101, and together with roh of 0.04, gives a design effect of 5. Still these design effects are larger than those normally obtained from a standard two-stage design in which sub-sampling within selected EAs reduces $\overline{B}$ and design effects.

A one-stage cluster design in which the selected EAs are completely enumerated is inefficient. It has large design effects which inflate standard errors. Moreover, this approach will not allow the CBS to produce tabulations of all information for new areas formed by an aggregation of adjacent EAs, such as new municipalities.

## 2.2 Sampling within every EA

This approach may appear to be expensive at first sight, but in the context of census enumeration it should be noted that since all EAs have to be visited for the 100 percent count, the cost of travelling to all EAs has already been incurred.

Two basic methods are discussed, the first involving one visit, while the second requires two visits. Each method has several variations.

(a) Method requiring one visit

In this approach the households are visited only once. The census enumerator is required to prepare the listing form, and at the same time interview the household, using either a short form, which contains the basic topics or a long form containing the basic questions and the supplementary questions. Those selected in the sample answer the long form while others answer the short form. To avoid selection biases, the short and long forms are arranged in a fixed sequence in an enumeration book. The sequence depends on the sampling fraction adopted and the type of form to be used for a household is determined by the order in which the enumerator visits the housing unit occupied by the household. In this way a systematic sample of housing units and households for the long form is obtained.

For a 20 percent sample, a set of questionnaire would contain four short forms followed by one long form. To start the systematic sample, a random number from 1 to 5 is selected for each EA. Depending on the selected number, the necessary forms in a set are crossed out and the enumerator is to start the interview with the first household in the EA with the first form that has not been crossed out. The proper application of the approach depends crucially on a systematic listing of the housing units in the EA, using a fixed starting point and systematic travel within the EA.

This approach is unlikely to be workable in Nepal, given the difficult terrain in many areas.

(b) Method requiring two visits

In this approach, two visits are made, the first to list the housing units, households and to obtain a preliminary population count. The listing sheet serves as the sampling frame for the EA and the sampled households answer both the basic and supplementary questionnaires, while non-selected households answer the basic questionnaire.

The enumerator is given the listing sheet which shows clearly after sample selection, which households should be interviewed using the basic questionnaire and which should be interviewed using both the basic and supplementary questionnaires. This approach is considered most feasible for the 2001 census of Nepal. To facilitate the work of the enumerators, the initial listing exercise, will be carried out by census supervisors, and will be undertaken some time before the actual census enumeration. The listing work

should include house numbering which involves either fixing a card containing a house number and other identification details or just writing the assigned house number on the housing unit. This will enable the census enumerator to locate the housing units during the second visit.

A variation of the above is to have the enumerators do the listing and to canvass all households in the EA using the basic questionnaire while doing the listing. The results are then returned to the supervisor who then samples the households, and a second visit is made to these households to obtain the supplementary information. The advantage of this variation is that the enumerator uses the same form in each visit, that is, the basic questionnaire on the first visit and the supplementary questionnaire on the second visit.

Another variation is to divide the supplementary topics into two approximately equal groups and to put each group on a different questionnaire. The list of housing units and households is randomly divided into two equal parts, and one half is enumerated on one questionnaire and the other half on the other questionnaire. Both questionnaires contain the basic topics in addition to half of the supplementary topics. While this variation has some attractive features, it results in a 50 percent sample for each set of sample questions. This may be more than is required. Moreover, cross tabulation of sample questions on one questionnaire with those of the other is not possible, since these questionnaires have been answered by different households.

After several discussions within CBS, the consensus is to make use of two visits, with listing, house numbering to be carried out by the census supervisor during the first visit. After this the housing units are sampled by the supervisor. The enumerator makes the second visit for census enumeration, using the basic questionnaire for every household in the EA and the supplementary questionnaire for the sampled households as determined by the census supervisor.

## 2.3 Basic and Supplementary Topics

At time of writing, the division of topics into basic and supplementary ones has not been finalised. However, CBS is proposing the following:

(a) Basic topics

1. Type of home occupied
2. Type of House ownership
3. Possession of agricultural holding and area
4. Ownership of land and area owned

5. Poultry and livestock
6. Small scale economic activities
7. Persons away in other countries
8. Caste/ethnicity
9. Sex
10. Age
11. Relationship to head
12. Religion
13. Language spoken
14. Citizenship

(b) Supplementary topics

1. Main source of drinking
2. Cooking fuel
3. Fuel for lighting
4. Type of toilet
5. Type of facilities
6. Deaths during last 24 months
7. Sex, age, date died and cause of death
8. Age
9. Place of births
10. Duration of stay
11. Place of residence 5 years ago
12. Literacy
13. Level of education
14. School attendance
15. Marital status
16. Age at first marriage
17. Number of children born
18. Number of live births last 24 months
19. Type of work last 12 months
20. Number of months worked during last 12 months
21. Usual occupation
22. Industry
23. Employment status
24. Reason for not working
25. Family situation

It is a clear that a large number of topics will be included in the 2001 census. In part, it is an attempt to provide sufficient data for analysis and study of gender issues presently under discussion.

In general, the inclusion of items in the 100 percent counts depends on the precision and amount of detail required for an item. An item is included if it is considered fundamental and needed with high precision, particularly for small areas. Items that are comparatively inexpensive to collect also tend to be included in the 100 percent counts.

For the purpose of sample design and preparing estimates from the sample , it is only necessary to ensure that the main items used as controls in tabulations are included in the 100 percent counts. From the list given, the most important are age and sex, and for some other tabulations, cast/ethnicity will be needed. For household tabulations, the topics on type of house and house ownership will be needed.

## 3 Main Features of Sample

Once the enumeration approach has been selected it is useful to give a description of the main features of the proposed sample. This is done under the following topics:

(a) Geographical coverage

The sample will cover all areas in Nepal.

(b) Scope

The 100 percent count will cover the total population, all households and housing units in the country. However, the sample will cover all private households. The institutional population such as military personnel, prisoners etc and the homeless will be providing information on the basic items only and will not be covered in the sample.

(c) Domain

Domains are sub groups of the population that have been identified in the planning stage to ensure that the sample will be able to provide information of a given precision. Since sub national areal units are most easily identified, these frequently form the domains.

The domains of the sample for the 2001 census are in the individual municipalities, and rural parts of each of the 75 administrative districts.

(d) Stratification

Domains are similar and frequently identical to strata if both are based on geographical areas. However the principal objective of stratification is to reduce sampling errors of estimates at the national level, especially if the strata created have low internal variability. Thus for the sample for the 2001 census, the 75 administrative districts form the main strata, with village Development Committees (VDCs) in rural areas and municipalities within the districts forming sub-strata.

The design also makes use of implicit stratification in addition to explicit stratification mentioned above. Systematic sampling of housing units and households within each EA, provides a sample of housing units and households in proportion to the different types of such units found in the EA, and thus gives a better spread of the units within the EA. Sometimes the list of housing units and households is reordered before systematic sampling in the hope of increasing the implicit stratification effect. This is likely to produce only a negligible effect. Moreover, reordering of the list will further complicate sample selection in the EA.

(e) Sampling frame and sampling unit

Sample selection will be carried out in each census EA. The sampling frame for each EA will be created during the initial listing operation. The sampling frame is, perhaps, the most important component of the overall sample design. It provides a means to identify and locate the sampling units. The listing form, among other items, will contain the serial number of the housing units and the serial number of households found within the housing units. This list frame of housing units forms the sampling frame for selecting the housing units, which constitutes the sampling unit, and all households and persons found in the selected units will be enumerated on both the basic and supplementary questionnaires. The sample of households and persons in each EA is thus a one-stage cluster sample, the cluster being the housing unit.

While a direct sample of households within the EA has greater precision, since there less clustering effect, interviewing, only some households and not others in the same housing unit can lead to queries and suspicion. Furthermore, if the listing operation is

carried out some time before census enumeration, selected households could have moved away. Selection of the housing unit and taking all households found at time of census enumeration avoid this problem. The clustering effect of sampling housing units is also likely to be small. From preliminary results of the 1981 census, on average, each housing unit contains only 1.1 households, which leads to a small deff.

(f) Sample Selection Procedure

Sample selection of housing units in each EA will be by systematic sampling. In principle systematic sampling is carried out by first computing the sampling interval I which is equal to 1/f, where f is the sampling fraction. For the present design I will always be a whole number. Systematic sampling involves the selection of sampling units at a fixed interval from a list, starting from a randomly determined point. The selection within each EA will be carried out by the census supervisor, who will be, on average, be in charge of 5 EAs.

Systematic sampling is generally simple to implement and errors in selection can be easily detected. This is the main reason why it is widely used in the field.

The Steps to be taken are summarised below.

a) Select a random number between 1 and the sampling interval from a random number table. The sampling interval will be supplied by CBS. For example, if the sampling interval is 8, then a random number will have to be selected from the numbers 1 to 8. For example, 4 is selected. This means that housing unit with serial number 4 in the listing form for the EA is selected, and information on the basic and supplementary items need be obtained from all households and persons found in housing unit no. 4.

b) The second step is to select the second housing unit. The second number to be selected is obtained by adding the first selected random number to the sampling interval, i.e. 4 + 8 = 12. Thus, housing unit with serial number 12 in the EA will be the second unit selected.

c) Step (b) is continued, and each time the sampling interval is added to the last number to select the next housing unit.

d) Selection is completed when the end of the listing form is reached.

e) Step (a) is again taken for the next EA. And this is followed by (b), (c) and (d).

Note should be taken that a new random number is to be selected for each EA.

# 4 Sample Size Determination

Initial discussions with CBS officials indicated that with given resources, a maximum sample for the country could be about 20 percent, while it would be useful to investigate whether a 10 percent sample would be sufficient. The sample size should also be adequate to provide gender specific data with precision sufficient for these data to be used in studies and policy formulation.

## 4.1 Overview of Sample Size Required

In this section we investigate what range of estimates can be produced satisfactorily, from a 20 percent and a 10 percent sample. We base our calculations on an average sized district of 300,000 persons in 2001 which is then expected to have 53,571 households. The method used in computing the results and assumptions used are given in Appendix 1.

Tables 1 & 2 show the approximate standard errors, relative error and 95 percent confidence interval for various subclasses within the district. The subclass is a group of persons possessing a certain characteristic, such as, the number of females, 25 and over with tertiary education. The estimates from a sample are subject to sampling errors which are indicated by the standard error. To facilitate comparisons, the relative error expresses the standard error as a percentage of the subclass. The last column shows the 95 percent confidence interval, which indicates that we are 95% confident that the actual (unknown) number of persons, is between the lower and upper limits given.

Table 1 shows results from a sample of 20 percent for a district with a population of 300,000. We see that estimates for subclasses below 100 persons are subject to relative errors of more than 20 percent. Though there is no fixed acceptable level of precision, for most practical purposes, an estimate with a relative error of more than 20 percent is frequently considered as insufficiently precise for many uses.

Table 1 : Approximate Standard Error, Relative Error And 95% Confidence Interval For Estimates Of Sub Classes For 20% Sample

**AVERAGE DISTRICT (300,000 persons in 2001)**

| Size Of Subclass | Standard Error | Relative Error (%) | 95% Confidence Interval | | |
|---|---|---|---|---|---|
| 50 | 14.1 | 28.3 | 22 | - | 78 |
| 100 | 20.0 | 20.0 | 60 | - | 140 |
| 250 | 31.6 | 12.6 | 187 | - | 313 |
| 500 | 44.7 | 8.9 | 411 | - | 589 |
| 1,000 | 63.1 | 6.3 | 874 | - | 1,126 |
| 2,500 | 99.6 | 4.0 | 2,301 | - | 2,699 |
| 5,000 | 140.2 | 2.8 | 4,720 | - | 5,280 |
| 10,000 | 196.6 | 2.0 | 9,607 | - | 10,393 |
| 25,000 | 302.8 | 1.2 | 24,394 | - | 25,606 |
| 50,000 | 408.2 | 0.8 | 49,184 | - | 50,816 |
| 100,000 | 516.4 | 0.5 | 98,967 | - | 101,033 |
| 200,000 | 516.4 | 0.3 | 198,967 | - | 201,033 |

We see that a 20 percent sample provides useable estimates for a subclass as small as 100 persons which constitutes 0.03 percent of the total. The 95% confidence interval tells us that we are 95 percent confident that the true, but unknown number, of this group is between of 60 and 140 persons.

A 20 percent sample of households will give, on average, a sample of 54,000 households. Table 1 can also be used to provide an idea of the expected precision of subclasses of households based on a 20 percent sample of households. For example, Table 1 shows that an estimate of a subclass, of 100 households is expected to have a relative error of 20 percent. But these 100 households represent a 0.2 percent of the household population.

The smaller sample size of 10,000 households account for the larger sampling errors obtained for household characteristics. For the rest of the report, considerations will be given in determining an appropriate sample size for households. This sample size will provide estimates of much higher precision for the population characteristics.

Table 2 shows the results of a 10 percent sample from an averaged sized district of population of 300,000. As expected, because of the smaller sample size, estimates from this sample have larger sampling errors and thus wider 95 percent of confidence intervals. A subclass of 100 persons is expected to have a relative error of 30 percent

and is considered not sufficiently precise for many users. Table 2 shows that the smallest subclass that can be estimated with sufficient precision is about 235 persons, forming 0.08 percent of the population.

Table 2 : Approximate Standard Error, Relative Error And 95% Confidence Interval For Estimates Of Sub Classes For 10% Sample

**AVERAGE DISTRICT (300,000 persons in 2001)**

| Size Of Subclass | Standard Error | Relative Error (%) | 95% Confidence Interval | | |
|---|---|---|---|---|---|
| 50 | 21.2 | 42.4 | 8 | - | 92 |
| 100 | 30.0 | 30.0 | 40 | - | 160 |
| 250 | 47.4 | 19.0 | 155 | - | 345 |
| 500 | 67.0 | 13.4 | 366 | - | 634 |
| 1,000 | 94.7 | 9.5 | 811 | - | 1,189 |
| 2,500 | 149.4 | 6.0 | 2,201 | - | 2,799 |
| 5,000 | 210.4 | 4.2 | 4,579 | - | 5,421 |
| 10,000 | 295.0 | 2.9 | 9,410 | - | 10,590 |
| 25,000 | 454.1 | 1.8 | 24,092 | - | 25,908 |
| 50,000 | 612.4 | 1.2 | 48,775 | - | 51,225 |
| 100,000 | 774.6 | 0.8 | 98,451 | - | 101,549 |
| 200,000 | 774.6 | 0.4 | 198,451 | - | 201,549 |

A 20 percent sample for all domains, resulting in a fixed sampling rate of 1 in 5 has two distinct advantages. It would be simple to implement, since all census supervisors would be trained to select the sample at a fixed rate and this design which is a self-weighting design, also provides the must precise estimates at the national level.

However, this design is not completely appropriate for Nepal, because the districts vary considerably in size, and there is even greater variability in size of urban places or municipalities. For example from population projections for 2001, Kathmandu, the largest district is expected to have 178,280 households, whereas a small district, such as Manang is expected to have only 1,320 households. A 20 percent sample for Manang will result in only 264 households.

A greater problem is encountered in sampling the municipalities. According to the 1991 Census, the population in municipalities only constituted 9.2 percent of the total population of Nepal. Current estimates show proportions range from a high of 68 percent in Kathmandu district and 45 percent in Lalitpur district to only 3 percent in Sarlahi district.

A sample based on one sampling rate is, thus, not practical for the census. It results in too small samples for some of the domains. The approach adopted is (1) to maintain a fixed sampling rate for most of the rural areas, (2) to make adjustments to this basic rate for municipalities and smaller districts to maintain sufficient sample size, and (3) to minimise the number of separate sampling fractions as much as possible to achieve operational simplicity. Basically, we need to reduce the sampling rate from 1 in 5 for the rural areas to reduce the number of cases, so that these can be used for the smaller districts and municipalities. For very small domains, a complete census would need to be taken. To achieve these objectives we need to determine a minimum domain sample size.

## 4.2 Determination of Minimum Domain Sample Size

While there is no single rule that can be used to determine a minimum sample size, we can proceed by taking into account two factors. The first is what is the smallest proportion of the domain population that may be required, and the second is what is the minimum level of precision acceptable for this proportion. For the household items, we expect the domain sample to provide estimates for a proportion of households of 1 percent to have a relative error of not more than 20 percent. Using standard notations, we can denote our requirement as follows.

Let p = 0.01, q= 1-p=0.99 and
R.e. (p) = 0.2

An initial sample size n' based on the above and a probability level of 68 percent is:
$n' = q /(p \times 0.2^2) = 2,475$

Since the domain population N is small, we adjust for finite population correction (1-f), where f is the sampling fraction, We obtain the required minimum sample size of households as $n = Nn'/N+n'$

The final sample size depends on the domain population, with a smaller sample size required for the smaller domain when we take into account (1-f). We take as a minimum a sample size of 2,000 households.

A sample of 2,000 households will produce, on average, 11,200 persons and for this size of sample we expect estimates as small as 0.5, percent of the population to have a

relative error of about 12 percent. Therefore, estimates of population characteristics will have much smaller sampling errors than household estimates.

### 4.3 Determination of Final Sample Size

The main guidelines used in deciding the final sample sizes for domains are: (1) the overall sample size should be between 20 percent and 10 percent., (2) the minimum sample size for households for a domain be about 2,000 households and (3) a minimum number of sampling fractions to be used to simplify sample selection in the field.

A sampling plan requiring four sampling rates, (a) 1 in 6; (b) 1 in 4, (c) 1 in 2 and (d) complete enumeration was then drawn up. In 15 of the districts, the same sampling fraction of 1 in 6 was to be used. For others, which contain both urban and rural areas, two separate rates were needed. For the two smallest districts, Mustang and Manang, complete enumeration using both the basic and supplementary questionnaires was required. The overall size was 18.2 percent of all households.

Internal discussions within CBS revealed a preference for a reduction in sampling rates and, if necessary, the total size could be increased to accommodate full enumeration in municipalities and smaller districts. The final results are given in Tables A1 and A2 in Appendices 2 and 3.

It can be seen that the domains where sampling will be carried out have a constant sampling fraction of 1 in 8, and thus the design is self-weighting. The majority of the rural districts, or those with a rural portion will be sampled, with the smallest six to be completely enumerated for both the basic and supplementary questionnaires. The design for the municipalities is quite different. Sampling will be used in six of the largest municipalities only. Given the preference of CBS for not using two different sampling fractions within the same district, complete enumeration has to be used for the rest of the 58 municipalities. The number of households that will be answering both the basic and supplementary questionnaires total 843, 842 which is 20 percent of all households.

## 5 Estimation Procedures

In this section we document the procedures for making estimates from the sample taken and the estimation of standard errors. While the estimation procedures are necessary for the preparation of census tabulations based on the sample results, the estimation of

standard errors is essential for assessing the reliability of the estimates. Moreover, these results are essential for statisticians in evaluating the design used and in finding ways to improve the design for the future.

## 5.1 Ratio Estimation

The ratio estimation method will be used in making estimates from the sample. It ensures that the sample estimates are generally consistent with the 100 percent counts and the estimates have smaller sampling errors. In general, the ratio estimation method can be shown as

$$y''_{hi} = \sum_j \frac{y_{hij}}{x_{hij}} X_{hij} \quad \text{..... (2)}$$

Where $y''_{hi}$ is the ratio estimator for the population with a certain characteristic in the ith domain and in the hth district. The number of persons found in the sample with a certain characteristic in the jth tabulation group, in the ith domain and in the hth district is $y_{hij}$. $x_{hij}$ is the total number of persons found in the sample in the jth tabulation group, in the ith domain and in the hth district. $X_{hij}$ is the total number of persons from the 100 percent count, found in the jth tabulation group, in the ith domain and in the hth district.

The estimator for the hth district is

$$y''_h = \sum_i y''_{hi}$$

where all domains have been sampled.

Where some domains are completely enumerated, the district estimator is

$$y''_h = \sum_i y''_h + Y_h$$

where $Y_h$ is the total from the completely enumerated domains.

The national level estimator is

$$y'' = \sum_h y''_h$$

Similarly, estimates for the five development regions and the three ecological zones can be obtained by adding together the district estimates falling in each of these areas.

## 5.2 Tabulation Groups

To implement the above, we need to develop the required tabulation groups. These groups are found by cross-tabulating selected items from the 100 percent count items. In selecting the items as control items, we take note of those items which form the main controls in the census tabulations and, preferably, the items in the sample should be correlated with the selected items. Often the correlation will be high if the sample items are subgroups of the control variables.

It is useful to discuss further the creation of the tabulation groups separately for tabulations for persons and those for households.

(a) Tabulations for persons

The main control variables for the majority of tabulations for persons will be age and sex. Thus the main tabulation groups must be formed from these two variables. The other consideration in forming the tabulation groups is that the expected number of samples cases in each of the cells created should not be too small.

The ratio estimate has a small technical bias, but this bias is negligible when the sample size is not too small. The coefficient of variation of the size of the numbers in each cell of the tabulation groups is used as a guide and this should, preferably, be less than 0.1.

Moreover, the estimation of standard errors of the ratio estimator is based on the Taylor Expansion or Delta Method which is appropriate only for large samples where the coefficient of variation of the sample sizes ($x_{hij}$) is less than 0.2 and preferably less than 0.1.

Using the age distribution of males and females in official projections for Nepal for the year 2001, and computing the expected number of cases in a small urban sample of 10,000 ( 2000 households) we obtain Table 3 shown below.

Table 3: Expected sample sizes by sex and age groups in a sample Of 10,800 persons

| Age Group | Male | Female | Total |
|---|---|---|---|
| 0-4 | 827 | 786 | 1613 |
| 5-9 | 749 | 708 | 1457 |
| 10-14 | 649 | 634 | 1283 |
| 15-19 | 526 | 560 | 1086 |
| 20-24 | 461 | 500 | 961 |
| 25-29 | 416 | 434 | 850 |
| 30-34 | 354 | 362 | 716 |
| 35-39 | 298 | 298 | 596 |
| 40-44 | 257 | 255 | 512 |
| 45-49 | 222 | 217 | 439 |
| 50-54 | 186 | 182 | 368 |
| 55-59 | 151 | 149 | 300 |
| 60-64 | 116 | 118 | 234 |
| 65-69 | 84 | 88 | 172 |
| 70+ | 100 | 114 | 214 |

The expected numbers are given for standard 5 year age groups and the group 70 years and over. It can be seen that even for a small sample of 10,000 persons the frequencies in each cell are sufficient to justify the use of the ratio estimation method.

At this stage it may be useful to illustrate with an example how formula (2) is to be used. Let us estimate the number of single females in the age group 20-24 who are living in Kathmandu municipality, and Kathmandu district. this number is obtained from the sample and is represented by $y_{hij}$. The corresponding denominator from the sample is the number in the age-sex tabulation group indicated by $x_{hij}$ which, in this case, is the number of females in the age group 20-24 in Kathmandu municipality. The corresponding value of $X_{hij}$ is the total number of females in the age group 20-24 in Kathmandu Municipality found in the 100 percent counts.

Thus the ratio $y_{hij}/x_{hij}$ represents in this example, the proportion of women who are single in the age group 20-24 for Kathmandu municipality, and this proportion is multiplied by the total number of females aged 20-24 in Kathmandu municipality from the 100 percent counts, $X_{hij}$ to give an estimate of the total number of single females

aged 20-24 in Kathmandu municipality. If similar estimates are made for the age groups from 10-14 to 70+, the sum of these estimates gives us the total number of single females in Kathmandu municipality.

## 5.3 Summary of Procedure

To implement the ratio estimation method outlined above for large-scale processing, it will be useful to first compute $X_{hij}/x_{hij}$ as weights, $\omega_{hij}$ .

Then the ratio estimate is computed by

$$y''_{hi} = \sum_j \omega_{hij} y_{hij}$$

The following steps are needed to compute the ratio estimates:

(1) Decide on the control variables from the 100 percent counts, and the maximum number of categories required, taking note that the expected frequencies in each cell are not less than 20 for a sample of 10,800 persons. This step gives the tabulation groups.

(2) Tabulate the 100 percent counts for the same tabulation groups in the domain.

(3) Tabulate the sample counts for the same tabulation groups in the domain.

(4) Compute the weights for the sample data by dividing the results in (2) by the results in (3) for the same tabulation groups.

(5) Enter the weights computed for each tabulation group on the person or household records.

(6) Tabulate the weighted sample totals, which are the required ratio estimates of the numbers with the required characteristic. For quantitative variables, such as the number of children ever born or age at first marriage, we multiply these numbers by the weight for each record and find the weighted sum to get the totals.

## 5.4 Other Points on Tabulations of Persons

(a) It will be necessary to have a break at age 6 years, for tabulation groups needed for tabulations on literacy and education attainment, since this information is obtained from persons 6 years and above.

(b) Some tabulations of persons require a control on caste/ethnic group which is available from the 100 percent counts. However since there are many different castes/ ethnic groups in the country, the tabulation groups from this variable will contain many small numbers. Tabulations at the national level using an caste/ethnic control will present no problems and, such tabulations can be made for the municipalities and districts which have a complete enumeration of both the basic and supplementary questionnaires.

(c) Tabulation groups need to be formed for household tabulations as well. Since the sample size for households is much smaller, these groups have to be fewer in number to maintain sufficient size in each group formed.

The main items from the 100 percent counts that are suitable for the formation of tabulation groups, given the types of household tabulations that will be prepared, are: (a) type of house occupied by household, (b) type of house ownership, (c) sex of household head and (d) age of household head.

We select the items, (a) type of house occupied and (b) type of house ownership to form the tabulation groups. These items are likely to be correlated with those collected for the household in the sample, such as (a) source of drinking water, (b) cooking fuel, (c) type of toilets, etc.

## 5.5 Conversion to Integral Weights

The computed weights $\omega_{hij}$ will in most cases, be fractional weights, such as 8.623, for example. If fractional weights are used in tabulations, it will be necessary to round the results before publication and there may be some inconsistency between tabulations.

To avoid this inconsistency due to rounding, the fractional weights should be converted to whole numbers. For example, if the fractional weights are computed to one decimal place, say 8.6, we give a weight of 9 to 6/10 of the group and 4/10 gets a weight of 8. Thus the average weight for the whole group is equal to $8.6\{(9\times{}^{6}\!/_{10})+(8\times{}^{4}\!/_{10})\}$.

A random process has to be used to determine which households or persons get which weights. Following the above example we select six random numbers between 0 and 9, for example, 1, 3, 4, 5, 8, and 2. For successive groups of 10 households and their

members in a tabulation group, we assign a weights of 9 to the first, second, third, fourth, fifth, and eighth households and their members. A weight of 8 is assigned to the remaining 4 households and their members.

## 5.6 Simple Expansion Method

The ratio estimation method is recommended for the preparation of all estimates from the sample. However, it involves fairly heavy computations. If computing capacity proves to be a major constraint, the simpler, but less efficient, simple expansion method may be used. This estimator is

$$y'_{hi} = F\sum_{j} y_{hij}$$

where $y'_{hi}$ is the simple expansion estimator for the number of persons or households with a certain characteristic in the ith domain and hth district. F is the reciprocal of the sampling fraction which is equal to 8 for all domains.

## 5.7 Estimation of Sampling Errors

The ratio estimates obtained from the supplementary sample are subject to sampling errors indicated by the standard error. An important feature of a probability sample, like the one that will be taken for the 2001 census, is that it is possible to provide estimates of standard errors directly from the sample results.

Since persons and households selected in the sample come from the sampled housing units where they were located, the most precise method for estimating the standard error requires computations to be carried out at the housing unit level. However this would prove to be impractical given the size of the sample involved. A method that requires less computations is required. We propose to use the random group method in which households and persons found in a sampled housing unit are assigned to groups at random. These groups are treated as separate subsamples or replicates and their estimates provide a way to estimate the variance of the average of the estimates from the subsamples. Below we set out the general formulas to compute the variances and standard error. Replicate sampling provides a simple way to estimate the variance. Given c estimates $\varphi_1, \varphi_2, ....., \varphi_c$ of a population parameter $\varphi$, obtained from independent replicates of the same design, the variance of the mean of the estimates $\overline{\varphi} = \sum \frac{\varphi_k}{c}$ is given by $\text{var}(\varphi) = \frac{\text{var}(\varphi_k)}{c}$

And var($\varphi_k$) is estimated from the c values as

$$\text{var}(\varphi_k) = \sum \frac{(\varphi_k - \overline{\varphi})^2}{c-1}$$

Thus, $\text{var}(\overline{\varphi}) = \sum \frac{(\varphi_k - \overline{\varphi})^2}{c(c-1)}$ ..... (3)

Formula (3) is the general formula suitable for estimating the variance of the average of the replicate estimates of the population, and it can be applied in the case where we randomly divide a total sample into random groups.

Assuming we divide the total sample for a domain into "c" random groups and we let $y_k''$ be the ratio estimate of the number with a specific characteristic in a domain from the kth random group[1].

An individual ratio estimate from one random group must be multiplied by c to provide an estimator of the domain total, since $y_k''$ is based on 1/c of the total sample. Thus, $\overline{\varphi} = \sum_{k=1}^{c} \frac{cy_k''}{c} = \sum_{k=1}^{c} y_k'' = y.''$ Substituting for $\overline{\varphi}$ and $\varphi_k$ in formula (3) we get

$$\text{var}(y'') = \frac{\sum_{k=1}^{c} (cy_k'' - y'')^2}{c(c-1)} \quad \text{..... ..... ..... (4)}$$

After simplification, and the inclusion of (1-f), because sample size is large, formula (4) can be shown to be

$$\text{var}(y'') = (1-f)\frac{\left[c\sum_{k=1}^{c}(y_k'')^2 - (y'')^2\right]}{c-1} \text{..... (5)}$$

The standard error of y" is

$s.e.(y'') = \sqrt{\text{var}(y'')}$ and the relative error (%) is

$$r.e.(y'')(\%) = \frac{s.e.(y'')}{y''} \times 100$$

To estimate the variance of proportions or, in general, ratios such as $r = \frac{y''}{x''}$,

we use the following formula:

$$\text{var}(r) = \frac{1}{(x'')^2}\left[\text{var}(y'') + r^2\,\text{var}(x'') - 2r\,\text{cov}(y'', x'')\right] \text{..... (6)}$$

where var(y") and var(x") are computed by formula (4) and the cov(y",x") is

[1] $y_k''$ is equivalent to $y_{hi}''$ in section 5.1, except it is for the kth random group.

$$(1-f)\frac{\left[c\sum_{k}^{c} y_k'' x_k'' - y''x''\right]}{c-1}$$

Estimates of variances at levels above the district, are obtained by simply summing up the variances at district level to the appropriate level required. Since sampled households and persons are clustered within housing units, it is useful to estimate the design effects. This is easily computed for proportions. The design effect is $deff = \frac{\text{var}(r)}{\text{var}(p)}$ where var(r) is given in (6) and $\text{var}(p) = (1-f)\frac{pq}{n-1}$ where p is the proportion estimated and is the same as r here and q=1-p. The number of either households or persons in the sample is indicated by n.

## 5.8 Implementation of the Random Group Method at Domain Level

(1) We need to determine the number of random groups c. In general, the number of random groups should be large enough to allow the variance estimator to have enough precision. At the same time, the sub-sample size for each group should not be too small. Given the large sample size that is used in the present case, 20 random groups can be formed to provide estimates of standard errors. Based on a minimum sample size of 2000 households for a domain, we expect 100 households per random group.

(2) We need to randomly allocate the households and person records within a selected housing unit to each of the 20 random groups within each domain.

The steps can best be shown separately for a municipality and the rural part of a district.

In a municipality, the census EAs are arranged by wards, and the wards are grouped according to neighborhoods within the municipality. the household and person records are arranged according to the serial number of their housing unit within an EA.

In the rural part of the district, the EAs, which in most cases, will be the wards, are arranged by VDCs, and similarly the household and person records are arranged according to their housing unit members within each EA/Ward.

(3) The details of the allocation of households and person records to the 20 random groups can then be implemented. For each domain, we select a random number between 01 and 20. For example, if 5 is selected for a domain, the household and person records

for the first housing unit in the sample will be given the random group number 5. Those households and persons in the second selected housing unit is assigned to group 6 and so on until the 20th unit is assigned to group 4. The household and person records for the 21st unit will be assigned to group 5 and we continue this process until the end of the list is reached. If the allocation of household and person records is carried out manually, then the random group number assigned has to be entered in the questionnaires before data input.

(4) The computation of standard errors based on random groups may be carried out as a semi-manual process. We first decide which key tabulations are required for standard error computation. At the time of tabulation, we arrange for these tabulations at domain level to be produced for each of the 20 random groups. These tabulations can be downloaded to a spreadsheet and the computations for the standard errors done with the spreadsheet.

**Appendix 1**

## 3.0 Method used in Computing Tables 1 and 2

As mentioned, the standard errors presented in Tables 1 and 2 are approximate and in this section we describe how they have been computed.

The proportion of the population possessing a given characteristic is estimated from a sample of size n as

$$p = \frac{m}{n}$$

where m is the number of persons with a given characteristics in the sample. The number of such persons in the population, is estimated by Np where N is the total population. This estimator is subjected to a sampling error, indicated by the standard error. The standard error is defined as the square root of the sampling variance, which in this case is:

$$\text{var}(Np) = N^2(1-f)\frac{pq}{n}$$

where q = 1-p and f = n/N is the sampling fraction. Since the sample size n is large we have used n instead of n-1 in (1).

The Standard error, s.e.($Np$) is defined as $\sqrt{\text{var}(Np)}$

and the Relative error (%) is $r.e.(Np)\% = \frac{s.e.(Np)}{Np} \times 100$

The 95% confidence interval is $Np \pm 2.0 \times s.e.(Np)$

The results presented are based on simplifying assumptions, but they should be accurate enough for use in making decisions on required sample sizes, since the factors not taken into account in the computations tend to balance each other. Sampling errors for the sample design finally adopted could be a little larger than those shown because of the clustering of persons in housing units. On the other hand, the use of systematic sampling introduces implicit stratification which will reduce sampling errors. Also, the ratio estimation method, to be used in making the final estimates will also lead to smaller sampling errors.

**Appendix 2**

## TABLE A1: EXPECTED NUMBER OF RURAL HOUSEHOLDS SAMPLED.

| | EST.POP. 2001 | RURAL HH 2001 | RURAL f | SAMPLE RURAL HH | TOT VDC WARDS | AVE.VDCHH PER WARD |
|---|---|---|---|---|---|---|
| **NEPAL** | **23453019** | **3660652** | | **489095** | **35217** | 14 |
| 5 Morang | 862537 | 131030 | 0.125 | 16379 | 585 | 28 |
| 4 Jhapa | 752865 | 118322 | 0.125 | 14790 | 423 | 35 |
| 17 Dhanusa | 695809 | 113508 | 0.125 | 14189 | 909 | 16 |
| 19 Sarlahi | 624366 | 108628 | 0.125 | 13579 | 891 | 15 |
| 15 Saptari | 588694 | 102827 | 0.125 | 12853 | 1026 | 13 |
| 16 Siraha | 582275 | 96377 | 0.125 | 12047 | 954 | 13 |
| 18 Mohattari | 552450 | 96264 | 0.125 | 12033 | 684 | 18 |
| 49 Rupendehi | 690553 | 93350 | 0.125 | 11669 | 621 | 19 |
| 48 Nawalparasi | 582833 | 92819 | 0.125 | 11602 | 657 | 18 |
| 32 Rautahat | 526132 | 92416 | 0.125 | 11552 | 864 | 13 |
| 33 Bara | 538276 | 85382 | 0.125 | 10673 | 882 | 12 |
| 6 Sunsari | 607948 | 83196 | 0.125 | 10399 | 441 | 24 |
| 50 Kapilbastu | 491592 | 76540 | 0.125 | 9567 | 693 | 14 |
| 71 Kailali | 588920 | 70300 | 0.125 | 8787 | 378 | 23 |
| 35 Chitwan | 467809 | 65523 | 0.125 | 8190 | 324 | 25 |
| 34 Parsa | 483369 | 64288 | 0.125 | 8036 | 738 | 11 |
| 30 Dhading | 345113 | 63635 | 0.125 | 7954 | 450 | 18 |
| 23 Sindhupalchowk | 316618 | 62215 | 0.125 | 7777 | 711 | 11 |
| 46 Gulmi | 323558 | 61404 | 0.125 | 7676 | 711 | 11 |
| 56 Dang | 461458 | 60800 | 0.125 | 7600 | 351 | 22 |
| 24 Kavre | 389166 | 59169 | 0.125 | 7396 | 783 | 9 |
| 31 Makawanpur | 405952 | 58933 | 0.125 | 7367 | 387 | 19 |
| 38 Tanahun | 335151 | 57632 | 0.125 | 7204 | 414 | 17 |
| 27 Kathmandu | 946568 | 56604 | 0.125 | 7076 | 513 | 14 |
| 39 Shyangja | 351490 | 56511 | 0.125 | 7064 | 540 | 13 |
| 57 Banke | 384959 | 55062 | 0.125 | 6883 | 414 | 17 |
| 36 Gorkha | 303759 | 54330 | 0.125 | 6791 | 594 | 11 |
| 28 Nuwakot | 303313 | 51844 | 0.125 | 6480 | 549 | 12 |
| 3 Ilam | 295144 | 49873 | 0.125 | 6234 | 432 | 14 |
| 45 Baglung | 278433 | 49619 | 0.125 | 6202 | 531 | 12 |
| 58 Bardiya | 392788 | 49090 | 0.125 | 6136 | 279 | 22 |
| 47 Palpa | 285345 | 47270 | 0.125 | 5909 | 585 | 10 |
| 13 Khotang | 252588 | 46997 | 0.125 | 5875 | 684 | 9 |
| 69 Achham | 235507 | 46265 | 0.125 | 5783 | 675 | 9 |
| 59 Surkhet | 295801 | 45789 | 0.125 | 5724 | 450 | 13 |
| 40 Kaski | 381244 | 44203 | 0.125 | 5525 | 387 | 14 |
| 14 Udayapur | 293656 | 44090 | 0.125 | 5511 | 396 | 14 |
| 10 Bhojpur | 233960 | 43616 | 0.125 | 5452 | 567 | 10 |
| 20 Sindhuli | 282075 | 42917 | 0.125 | 5365 | 477 | 11 |
| 21 Ramechhap | 231372 | 42772 | 0.125 | 5347 | 495 | 11 |
| 51 Arghakhachi | 222255 | 42404 | 0.125 | 5301 | 378 | 14 |
| 72 Kanchanpur | 354866 | 41528 | 0.125 | 5191 | 171 | 30 |
| 52 Pyuthan | 212394 | 40335 | 0.125 | 5042 | 441 | 11 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 53 | Rolpa | 213785 | 39360 | 0.125 | 4920 | 459 | 11 |
| 22 | Dolakha | 214364 | 39331 | 0.125 | 4916 | 459 | 11 |
| 74 | Baitadi | 243468 | 38680 | 0.125 | 4835 | 558 | 9 |
| 55 | Salyan | 221497 | 38564 | 0.125 | 4821 | 423 | 11 |
| 2 | Panchthar | 214798 | 38559 | 0.125 | 4820 | 369 | 13 |
| 60 | Dailekh | 228388 | 37799 | 0.125 | 4725 | 495 | 10 |
| 54 | Rukum | 192497 | 36246 | 0.125 | 4531 | 387 | 12 |
| 37 | Lamjung | 178799 | 35550 | 0.125 | 4444 | 549 | 8 |
| 70 | Doti | 200704 | 35090 | 0.125 | 4386 | 450 | 10 |
| 44 | Parbat | 174357 | 33977 | 0.125 | 4247 | 495 | 9 |
| 25 | Lalitpur | 331212 | 32275 | 0.125 | 4034 | 369 | 11 |
| 12 | Okhaldhunga | 162923 | 30798 | 0.125 | 3850 | 504 | 8 |
| 68 | Bajhang | 168789 | 30447 | 0.125 | 3806 | 423 | 9 |
| 7 | Dhankuta | 178714 | 29041 | 0.125 | 3630 | 315 | 12 |
| 9 | Sankhuwasava | 171246 | 27916 | 0.125 | 3490 | 297 | 12 |
| 61 | Jajarkot | 139680 | 25540 | 0.125 | 3193 | 270 | 12 |
| 1 | Taplejung | 139153 | 24770 | 0.125 | 3096 | 450 | 7 |
| 43 | Magdi | 118454 | 24381 | 0.125 | 3048 | 360 | 8 |
| 11 | Solukhumbu | 117517 | 23252 | 0.125 | 2906 | 306 | 9 |
| 8 | Tehrathum | 124888 | 22313 | 0.125 | 2789 | 288 | 10 |
| 75 | Darchula | 123639 | 21421 | 0.125 | 2678 | 369 | 7 |
| 67 | Bajura | 111785 | 21312 | 0.125 | 2664 | 243 | 11 |
| 64 | Kalikot | 107526 | 19423 | 0.125 | 2428 | 270 | 9 |
| 73 | Dadeldhura | 130771 | 19256 | 0.125 | 2407 | 180 | 13 |
| 26 | Bhaktapur | 216328 | 17311 | 0.125 | 2164 | 144 | 15 |
| 63 | Jumla | 91653 | 16345 | 0.125 | 2043 | 270 | 8 |
| 29 | Rasuwa | 46233 | 9053 | 1 | 9053 | 162 | 56 |
| 65 | Mugu | 42647 | 8018 | 1 | 8018 | 216 | 37 |
| 66 | Humla | 42881 | 7682 | 1 | 7682 | 243 | 32 |
| 62 | Dolpa | 30519 | 6062 | 1 | 6062 | 207 | 29 |
| 42 | Mustang | 17279 | 3880 | 1 | 3880 | 144 | 27 |
| 41 | Manang | 5566 | 1320 | 1 | 1320 | 108 | 12 |

**Appendix 3**

**TABLE A2: EXPECTED NUMBER OF HOUSEHOLDS SAMPLED IN MUNICIPALITIES**

| NAME OF DISTRICTS | MUNICIPALITY | HOUSEHOLD 2001 | URBAN f | SAMPLE URBAN | NO. OF WARDS | HH PER WARD |
|---|---|---|---|---|---|---|
| | **TOTAL URBAN** | **558516** | | **354747** | | |
| Ilam | Ilam | 3500 | 1 | 3500 | 9 | 389 |
| Jhapa | Bhadrapur | 3627 | 1 | 3627 | 15 | 242 |
| | Damak | 9693 | 1 | 9693 | 19 | 510 |
| | Mechi | 8974 | 1 | 8974 | 13 | 690 |
| Sankhuwasava | Khadhbari | 4548 | 1 | 4548 | 13 | 350 |
| Dhankuta | Dhankuta | 4440 | 1 | 4440 | 9 | 493 |
| Morang | Biratnagar | 30731 | 0.125 | 3841 | 22 | 175 |
| Sunsari | Inarua | 4436 | 1 | 4436 | 10 | 444 |
| | Ithari | 6736 | 1 | 6736 | 9 | 748 |
| | Dharan | 16460 | 0.125 | 2058 | 19 | 108 |
| Udaypur | Triyuga | 9755 | 1 | 9755 | 17 | 574 |
| Saptari | Rajbiraj | 5540 | 1 | 5540 | 10 | 554 |
| Siraha | Siraha | 4843 | 1 | 4843 | 9 | 538 |
| | Lahan | 4577 | 1 | 4577 | 10 | 458 |
| Dolakha | Bhimeshwor | 5045 | 1 | 5045 | 13 | 388 |
| Sindhuli | Kamalamai | 5630 | 1 | 5630 | 18 | 313 |
| Dhanusa | Janakpur | 12373 | 1 | 12373 | 16 | 773 |
| Mahottari | Jaleshwor | 3696 | 1 | 3696 | 13 | 284 |
| Sarlahi | Malangawa | 3045 | 1 | 3045 | 10 | 305 |
| Nuwakot | Bidur | 4620 | 1 | 4620 | 11 | 420 |
| Lalitpur | Lalitpur | 26578 | 0.125 | 3322 | 22 | 151 |
| Kathmandu | Kathmandu | 113726 | 0.125 | 14216 | 35 | 406 |
| | Kirtipur | 7950 | 1 | 7950 | 19 | 418 |
| Bhaktapur | Bhaktapur | 11491 | 1 | 11491 | 17 | 676 |
| | Madhyapur Thimi | 6420 | 1 | 6420 | 17 | 378 |
| Kavreplanchowk | Dhulikhell | 1949 | 1 | 1949 | 9 | 217 |
| | Banepa | 2347 | 1 | 2347 | 11 | 213 |
| | Panauti | 4490 | 1 | 4490 | 13 | 345 |
| Makawanpur | Hetauda | 13446 | 1 | 13446 | 11 | 1222 |
| Rautahat | Gaur | 4445 | 1 | 4445 | 13 | 342 |
| Bara | Kalaiya | 3897 | 1 | 3897 | 14 | 278 |
| Parsa | Birgunj | 19000 | 0.125 | 2375 | 19 | 125 |
| Chitwan | Bharatpur | 14408 | 1 | 14408 | 14 | 1029 |
| | Ratnanagar | 6042 | 1 | 6042 | 13 | 465 |
| Gorkha | Prithvinarayan | 4986 | 1 | 4986 | 11 | 453 |
| Tanahun | Byas | 4636 | 1 | 4636 | 11 | 421 |
| Kaski | Lekhnath | 8023 | 1 | 8023 | 15 | 535 |
| | Pokhara | 26384 | 0.125 | 3298 | 18 | 183 |
| Syanja | Putalibazar | 6179 | 1 | 6179 | 13 | 475 |
| | Waling | 3766 | 1 | 3766 | 11 | 342 |
| Baglung | Kalika | 3521 | 1 | 3521 | 11 | 320 |
| Palpa | Tansen | 3259 | 1 | 3259 | 15 | 217 |
| Nawalparasi | Ramgram | 4135 | 1 | 4135 | 13 | 318 |
| Rupandehi | Butwal | 12160 | 1 | 12160 | 15 | 811 |
| | Sidhartha | 9086 | 1 | 9086 | 14 | 649 |
| Kapilbastu | Kapilbastu | 4050 | 1 | 4050 | 14 | 289 |
| Dang | Tulsipur | 5427 | 1 | 5427 | 11 | 493 |
| | Tribhubannagar | 6816 | 1 | 6816 | 11 | 620 |
| Dailekha | Narayan | 3543 | 1 | 3543 | 9 | 394 |
| Banke | Nepalgunj | 11096 | 1 | 11096 | 17 | 653 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Bardiya | Gularia | 6785 | 1 | 6785 | 14 | 485 |
| Surkhet | Birendranagar | 6254 | 1 | 6254 | 12 | 521 |
| Doti | Dipayal | 2797 | 1 | 2797 | 14 | 200 |
| Kailali | Dhanagadhi | 10203 | 1 | 10203 | 14 | 729 |
| | Tikapur | 5361 | 1 | 5361 | 9 | 596 |
| Kanchapur | Mahendranagar | 13588 | 1 | 13588 | 19 | 715 |
| Dadeldhura | Amargadhi | 3864 | 1 | 3864 | 11 | 351 |
| Baitadi | Dhasarathchand | 4139 | 1 | 4139 | 13 | 318 |