Demography India Vol. 28, No. 2 (1999), pp. 167-178

M. Mazharul Islam*

Small Area Estimation Methods with Application to Contraceptive Prevalence Rates in Bangladesh

Introduction

IN recent years, the demand for small area (such as counties or districts or municipalities or other small domain) statistics has greatly increased. This is due, among other things, to their growing use in formulating policies and programs, in the dispensation of government funds and in regional planning. Legislative acts by national governments have also increasingly created a need for small area statistics, and this trend will most likely continue. As a consequence, the generation of reliable small area statistics has emerged as a pressing and frequently difficult and costly problem. Estimates for local areas and other small domains have been of general interest for a long time, but have not been available except for estimates from population censuses, special surveys or administrative registers (Platek *et al.*, 1987). However, in recent years, program managers and policy makers in developing countries have expressed interest in obtaining estimates of basic demographic indicators for small areas in order to set targets, allocate resources and for the planning of reforms, welfare and administration in many fields, including health programs because at this level development planning takes place (Muhuri and Rutstein, 1994). Project and program activities are designed with local activities in mind.

The common sources of small area data are the various censuses of population, housing and agriculture. But these censuses are usually relatively infrequent, slow to produce results, and tend to concentrate on a comparatively small number of major variables of interest. Thus, the censuses, although providing a vital source of regular benchmark data, cannot meet the need for more frequent and more detailed information

'Associate Professor, Department of Statistics. University of Dhaka, Dhaka-1000, Bangladesh.

for subgroups, which is a feature of present-day administration. So, effective planning of health services and other governmental activities cannot depend on these traditional data sources; the data must be more current and more complete than these sources provide.

On the other hand, the use of service statistics or administrative records often has serious limitations for a variety of reasons. Conducting nationally representative surveys (such as DHS) in countries that include geographical areas containing small population is possible. However, such surveys rarely provide enough data to permit accurate estimation for small areas, because national level surveys are usually designed for national and at best for some large regional level estimates. Even if such estimates for sub-regional level are possible, they are often unreliable for small areas because the samples are of insufficient size to be representative and they may be suffering from large sampling errors. In addition, non-sampling errors may increase with increased sample size.

Thus, in practice, facing with impossibility of obtaining reasonable measures of demographic parameters for small area directly from the traditional data sources, some indirect methods have been developed to generates estimates for small population area (say districts).

The objectives of this study is to provide an overview of two selected indirect methods of small area estimation - the *synthetic* estimation procedure and the *regression— based* procedure. The application of these methods have been illustrated for estimating Contraceptive Prevalence Rates (CPR) of 60 districts (which we defined as small area) of Bangladesh. The results are then compared with the direct estimation procedure.

The Methodology

This study utilizes data extracted from the 1993-94 Bangladesh Demographic and Health Survey (BDHS). Some auxiliary information were also obtained from the 1991 census. The BDHS is part of the world-wide Demographic and Health Surveys (DHS) program, which is designed to collect data on fertility, family planning, and maternal and child health. The details of the survey may be seen elsewhere (Mitra *et al*.. 1994).

In the BDHS 1993-94, a total of 9,640 ever-married women aged 10 to 49 were surveyed individually. Among them 8989 were currently married. Bangladesh was divided into five administrative divisions, 64 districts (zillas) and 489 thanas. In rural areas, thanas are divided into unions and then mauzas, an administrative land unit. The primary sampling unit (PSU) was the mauza for rural areas. Urban areas are divided into wards and then mahallas. For urban areas, the PSU was mahalla. In total 304 PSUs were selected with probability proportional to size (PPS) sampling. Field work in three sample points was not possible. So a total of 301 points were covered in the survey.

Since BDHS 1993-94 was nationally representative, all the districts have relatively small population and, therefore, have a corresponding sample size that is too small to provide accurate estimates of demographic parameters. The only exception is Dhaka district in Dhaka division. So, in this analysis, the districts are considered as small areas. A cluster/PSU is the smallest unit of each district. The number of selected clusters with in district varies from 1 to 23. Ultimately we were able to obtain the estimate for 60 districts out of 64 districts of Bangladesh. Two districts, Bandarban of Chittagong division and Meherpur of Khulna division were not covered in the BDHS and the data required for estimate were not available for two other districts (Rangamati and Khagrachhari) of Chittagong division.

We first obtained direct estimates of the CPR and then compared them with the indirect estimates. Direct estimation of contraceptive prevalence rate (CPR) is defined as the ratio of the weighted number of women who are using contraceptive methods in an area to the weighted number of currently married women of age 10-49. The formula for direct estimation of the contraceptive prevalence rate, *R*, is as follows:

$$R = \frac{\sum_{i}^{n_i} W_i Y_i}{\sum_{i}^{n_i} W_i X_i}$$

where,

R = the contraceptive prevalence rate in a given administrative area (i.e. district);

Wi = the sampling weight for the *i*th woman;

Yi = whether the *i*th woman is using contraception currently (1 for yes and 0 for no);

Xi = the counting variable whose value is 1 for every currently married woman;

n_i; = the number of currently married women in the district.

Each direct estimate *R* is statistically unbiased for the parameter under study (i.e. CPR) in the small area (i.e. district), but it can be affected by large sampling error (SE) because of the relatively small sample size in the district. The sampling error can be used to calculate confidence intervals within which the true value for the population can reasonably be assumed to fall.

Using the direct procedure, the contraceptive prevalence rate (CPR) was estimated for 5 divisions and 60 districts of Bangladesh. Sampling errors and relative errors were not calculated for Madaripur in Dhaka division, Narail and Magura in Khulna division, Panchagarh, Thakurgaon and Lalmonirhat in Rajshahi division. Each of these districts have fewer than 3 PSUs and, consequently insufficient degrees of freedom for calculation of sampling error.

Since relative error is an important measure of accuracy (the smaller the relative error, the better the estimate), the districts that have smaller relative errors produce better estimates. The relative errors range from 1.3 to 18.3 percent (Table 1). In our study, all the districts have fewer than 25 PSUs and hence have relative errors greater than 5 percent (the larger the number of PSUs in the area, the better the estimate). The exception is Netrakona and Sherpur in Dhaka division and Khulna in Khulna division with relative errors less than 4 percent.

Synthetic Estimation Procedure

One of the limitations of the direct method is that the estimate is suffered by the large sampling error if the sample size is small. To overcome this problem, an alternative method of producing small area estimates is the population of synthetic estimation methodology. Synthetic estimation involves calculating estimates for larger areas (e.g. division, zones or domain) so that the sample size is sufficient to ensure their accuracy;
these estimates are then used to derive estimates for smaller sub-areas on the assumption that the smaller areas have the same characteristics as the larger area (Gonzalez, 1973). Synthetic estimation utilizes auxiliary data, for example on the distribution of respondents by age or educational attainment, that are correlated with the variable under study, for example contraceptive use. The population is first divided into sub-groups according to the auxiliary variable. Information about the relationship between the auxiliary variable and the study variable is obtained from the sample at the larger (division, zone) level next, an estimate is obtained for each sub-group in the sub-area by weighting the sub-group estimates across all categories of the auxiliary variable to produce the synthetic estimate for the sub-area.

The method of synthetic estimation rests on the assumption that the relationship between the auxiliary and study variables at the larger area level also holds for each sub-area. If true, synthetic estimation can produce unbiased estimates for sub-areas by borrowing from larger areas; however, if the assumption does not hold, the resulting synthetic estimates will be biased.

Depending upon the source of auxiliary information, the synthetic estimation technique is classified as: (i) Indirect and (ii) Standard.

In indirect synthetic estimation the auxiliary information are obtained from the sample itself. On the other hand when the auxiliary information are obtained from some external source such as census, the procedure is termed as standard synthetic estimation.

In this study we test the suitability of the synthetic estimation for obtaining small area estimates of CPR considering two auxiliary variable namely age and education. Age and education are two important co-variates of contraceptive use. Previous studies have shown that age and contraceptive use rate have curvilinear relationship. Contraceptive use rate increases with an increase in women's age and reach to the peak at the age group 30-39 and then decreases. Logistic regression analysis also shows that the log odds of contraceptive use are higher for women who are age 30-39 than for other age group. On the other hand contraceptive use has positive relationship with education level.

Indirect Synthetic Estimation Procedure

In this estimation procedure the estimate at the small area level is to be calculated by scaling down the large area estimate with the corresponding sample distribution of the auxiliary variable at the small area level without auxiliary external information. The assumption is that the characteristics of the sub-groups of the auxiliary variable at the small area level are the same as at the large area level. The procedure can be expressed as follows :

$$Pi = \Sigma_j(w_{ij} \times p_{.j})$$

where,

i represents a given sub-area (i.e. district) in the division,
j represents a given category for the auxiliary variable (i.e. age/education),
$p_{.j}$ represents the prevalence estimate for they'th auxiliary category at division level,
$w_{ij} = n_{ij} / n_i$ is the proportion of the sample population for the *j*th auxiliary category in the rth district, and,
p, represents the indirect prevalence estimate for the *ith* district.

Indirect synthetic estimation procedure is helpful when none or very little auxiliary information (from censuses) is available for forming weighting factors and as a consequence the room for choice of the adjustment weighting factors is very limited.

We have used two separate auxiliary variables—age and education of respondent as weighting factors in the estimation of the prevalence rate for small areas (districts) which are presented in Table 1.

Barguna in Barisal division is used as an example to show how the calculation is performed. The adjustment weights (w,.), for auxiliary variable age are .177, .463, .231 and .130 for age groups 10-19, 20-29, 30-39 and 40-49, respectively; the division level contraceptive prevalence rates are 30.7, 46.6, 57.7 and 47.2 respectively, for the four age groups. Using the above formula, we estimate the contraceptive prevalence rate 46.5 percent for Barguna district (shown in Table I): (.177 x 30.7)+(.463 x 46.6)+(.231 x 57.7) + (.130 x 47.2) = 46.5. Similarly, using education as auxiliary variable, the CPR for Barguna district has been obtained as 48.3 percent.

In this calculation, the assumption is made that the districts within the divisions are homogeneous with respect to age-specific or education-specific contraceptive prevalence rates. Even if the assumption of homogeneity is not correct, any possible extreme estimate is brought in line with the regional estimate.

The results indicate that the relative errors of indirect synthetic estimates using both age and education as auxiliary variables are smaller than that of direct estimates. That is, the indirect synthetic estimation works sufficiently well for all of the districts since it reduces relative errors of the estimates. But the procedure has some drawbacks. One

TABLE 1: SMALL AREA ESTIMATES OF CPR BY DIFFERENT METHODS: BANGLADESH, 1993-94

Division and Districts	*Direct Method*		*Indirect Synthetic estimation procedure*				*Standard Synthetic estimation procedure*		*Regression Method*	
			Age as auxiliary variable		*Education as auxiliary variable*		*Age as auxiliary variable*	*Education as auxiliary variable*		
	CPR	*RE*	*CPR*	*RE*	*CPR*	*RE*	*CPR*	*CPR*	*CPR*	*RE*
Barisal										
Barguna	52.9	7.4	46.5	2.6	48.3	.8	47.7	45.6	48.8	2.5
Jhalakati	42.4	5.7	47.5	2.3	48.9	2.7	48.3	46.2	47.2	2.5
Barisal	44.0	11.1	48.5	2.3	46.5	3.6	48.4	46.1	47.7	5.0
Bhola	56.8	5.5	48.1	7.2	49.0	6.7	48.1	43.3	50.8	9.8
Patuakhali	38.6	6.0	48.7	3.5	44.8	1.8	47.9	45.0	43.5	3.6
Pirojpur	50.2	7.6	46.0	3.5	49.5	2.4	48.3	46.9	47.6	6.5
Rajshahi										
Rajshahi	61.9	8.1	54.2	5.7	54.6	5.5	55.3	54.7	44.9	5.3
Dinajpur	59.2	12.7	55.8	8.2	55.4	6.0	55.5	54.8	47.7	3.1
Nilphamari	39.3	18.3	53.8	1.9	54.3	.6	55.7	54.2	43.0	3.7
Rangpur	57.9	9.2	54.7	2.2	54.4	8.8	55.6	54.5	47.7	8.8
Bogra	59.0	7.3	55.7	3.4	55.2	1.6	55.6	54.7	45.5	6.4
Naogaon	60.0	6.2	54.4	3.1	55.1	1.5	55.0	54.5	51.2	4.7
Nawabgonj	45.9	5.7	53.5	2.2	55.5	.7	54.5	54.2	43.8	8.9
Serajgonj	59.7	8.5	54.6	3.6	55.5	9.2	56.1	54.2	47.9	8.6
Pabna	50.9	17.3	56.2	1.6	54.2	4.1	55.7	54.6	44.6	7.2
Panchagarh	51.4	a	54.7	a	55.1	a	56.1	54.0	43.2	a
Thakurgaon	48.6	a	51.7	a	54.6	a	56.2	54.3	38.6	a
Lalmonirhat	53.6	a	58.4	a	54.7	a	55.7	54.1	43.0	a
Kurigram	46.8	9.6	55.9	a	54.3	a	55.3	54.1	39.8	3.0
Gaibandha	57.2	5.8	54.1	1.4	55.1	4.2	55.5	54.4	43.3	4.8
Joypurhat	50.0	a	54.2	2.7	55.3	.6	55.1	54.7	46.9	a
Natore	63.1	5.4	54.8	.7	54.5	.6	55.2	54.6	42.7	3.3
Khulna										
Khulna	62.7	2.5	55.8	3.4	55.7	5.0	56.1	55.6	49.9	5.2
Kustia	60.8	3.1	55.9	3.9	54.8	1.6	55.2	55.2	39.3	7.1

Chuadanga	44.8	3.8	52.1	.8	55.0	.7	54.4	54.9	39.8	4.5
Jhenaidah	67.2	8.6	55.0	3.3	55.6	1.3	55.0	54.9	43.9	10.3
Narail	64.8	a	56.1	a	56.1	a	56.1	55.5	52.4	a
Jessore	53.4	3.5	55.8	2.5	55.0	2.2	55.3	55.3	42.4	7.3
Bagerhat	48.7	3.2	55.7	2.5	55.3	.7	56.2	55.4	46.0	10.4
Magura	51.2	a	56.2	a	56.1	a	55.9	55.2	52.8	a
Satkhira	41.8	4.3	54.6	1.5	54.8	.5	55.1	55.1	41.5	5.5
Chittagong										
Chittagong	33.0	4.3	28.8	6.3	31.2	7.4	30.4	30.1	49.4	3.8
Sunamganj	15.7	10.4	27.7	6.9	27.8	16.5	30.7	27.5	37.0	5.7
Maulvibazar	18.1	3.4	29.0	4.1	29.3	6.5	30.6	28.0	38.4	16.1
Comilla	36.8	4.1	29.3	9.2	28.7	9.1	29.8	28.7	42.6	4.0
Chandpur	32.0	9.4	28.3	2.8	29.5	2.0	30.0	29.2	41.7	4.8
Lakshmipur	25.7	8.9	28.9	2.4	29.2	2.7	29.6	28.7	38.1	6.0
Feni	37.8	6.5	28.9	8.7	30.2	6.3	30.1	29.5	47.3	9.9
Syllet	18.9	7.3	30.4	2.0	27.8	1.8	30.6	28.4	41.3	4.1
Habiganj	23.0	7.2	30.7	3.3	27.4	2.6	30.3	27.9	42.9	5.4
Brahmanbaria	30.5	10.3	30.3	3.0	28.9	4.5	30.4	31.2	43.0	10.2
Noakhali	35.0	7.9	29.8	3.0	29.6	2.7	29.7	28.9	41.1	6.8
Cox's Bazar	33.3	4.4	27.5	3.3	28.4	1.8	29.6	27.7	41.9	3.3
Dhaka										
Dhaka	56.1	2.9	48.6	3.7	48.9	11.0	45.9	44.2	52.1	3.8
Narayanganj	51.8	9.0	44.4	.9	44.9	3.8	45.8	43.6	51.1	4.3
Gazipur	50.8	3.4	45.2	6.4	44.8	2.7	44.9	42.8	49.8	3.4
Faridpur	38.4	15.7	46.7	.6	43.7	8.2	45.6	42.6	42.5	14.8
Rajbari	50.0	5.3	45.4	6.8	44.2	12.0	45.0	41.9	44.0	19.5
Narsinghdi	28.7	7.4	43.6	9.3	43.0	10.9	45.7	43.1	46.5	9.5
Tangail	53.2	7.5	44.3	5.4	42.6	6.5	45.0	42.1	41.1	6.8
Jamalpur	54.2	3.0	42.3	6.8	44.1	1.8	45.1	41.8	44.6	4.7
Mymensingh	39.5	3.6	44.4	5.6	43.3	6.0	45.1	42.3	43.2	3.2
Kishoreganj	32.1	2.9	43.9	2.5	42.3	14.1	45.7	41.9	38.2	7.6
Netrakona	47.1	1.4	45.0	.9	43.6	2.8	45.9	41.9	45.4	7.5
Gopalganj	31.4	3.7	43.9	2.5	45.0	4.4	46.6	43.2	42.8	.7
Madaripur	54.1	a	44.8	a	45.4	a	45.7	42.4	46.7	a
Shariatpur	23.5	7.7	43.5	4.1	41.3	.5	46.2	41.6	35.9	2.5
Manikganj	44.0	8.7	41.0	4.1	41.6	1.0	45.5	42.6	40.3	8.4
Munshiganj	34.7	6.0	45.4	3.3	42.5	.9	46.5	43.3	59.9	4.7
Sherpur	40.4	1.3	43.1	1.2	40.9	1.2	45.1	41.2	40.5	8.9

of the problem with indirect synthetic estimation technique is that since it uses information from only the sample itself, the estimates can be affected by sample size variations. Standard Synthetic Estimation procedure can be used as a solution to this problem. In addition, in indirect synthetic estimation procedure, the assumption of homogeneity is required which may not always exists.

Standard Synthetic Estimation Procedure

The standard synthetic estimation is structurally similar to indirect synthetic estimation, with the major difference being in how the adjustment weights are calculated. In indirect synthetic estimation procedure, the weights are calculated from the sample itself while the standard synthetic estimation procedure makes use of some alternative external information either from an available recent census or from some other survey with a bigger sample for its weighting factors, assuming the variable of interest has been collected in the same source. For the districts of Bangladesh, the weights were calculated using the information from the 1991 Population Census, that is

$$w_{ij} = \frac{N_{ij}}{N_i}$$

where,

N_{ij} represents the estimated target population in the jth auxiliary category for the *i*th district from the 1991 population census, and,

N_j represents the estimated target population for the ith district in the 1991 population census.

In order to overcome the sample size variations of the indirect synthetic estimates, we have applied the standard synthetic estimation to the same data using woman's age and education as two weighting factor in the estimation of the prevalence rate for districts. The other calculations are performed as before and the results are presented in Table 1.

The results presented in Table 1 show that the estimates obtained by the standard synthetic estimation procedure show a close estimate of CPR obtained by the indirect synthetic estimation procedure but there is generally less variation between district level estimates within the divisions for the estimates obtained by indirect synthetic estimation. Both age and education as auxiliary variable produced similar results. Because of the fact that the sampling frame for the auxiliary age and education variables are different from that of the contraceptive use variable, the sampling variance of the standard synthetic estimate has not been calculated.

Regression Method

The accuracy of synthetic estimates discussed in the earlier sections largely depends on the choice of the auxiliary variable which correlates with the variable under study. One of the difficulties is that unless the auxiliary variable is highly correlated with the variable of interest, the synthetic estimates will tend to cluster near the mean for large domain and fail to reflect the actual effects of local area factors. Secondly, synthetic estimates are biased estimates. For small areas, bias occurs because the estimated rates for the characteristic of interest for the larger area, differ from the sub-area rates for most of the sub-areas. The change in population structure of past census data introduce an additional bias for synthetic estimates. One of the main weakness of the synthetic method is that only one auxiliary variable can be used at a time. Another problem is that the sampling and non-sampling errors of the auxiliary variables from the external source are not compatible with those of the estimates from the main source.

In order to alleviate these problems, *regression method* has been suggested which considers all the potential predictors related to the variable under study, while relying on data from the main source. The regression method is based on a regression equation using proportions of selecting indicator/predictor variables, measured for each cluster (smallest unit of a district), as independent variables and relies heavily on these proportions to determine the parameters in the equation. The multivariate logistic regression analysis suggest that use of contraception is significantly associated with a women's age, education, number of living children, work status, urban-rural residence and husband's education. These 6 variables are reformulated into 6 dummy variables (1,0) based on the cutoffs determined through a two-step process to maximize the power of discrimination of the predictor (Goldstein and Dillon, 1978). At the first step, the cumulative distribution of each of the predictors is constructed for two subgroups : those who are currently using a family planning method and those who are not. Then, the absolute difference in the cumulative frequency between the two subgroups is calculated for each of the categories of the predictor. The cutoff point, x^*, is that value (or the category) of the original predictor which maximizes the difference in the cumulative frequency between use and non-use subgroups, which can be expressed as follows :

$$| Fu(x^*) - F_{,,}(x^*)\} = \max_x | F_,(x) - Fn(l) |$$

where,

Fu (x) = the cumulative frequency distribution of the original predictor at value *x* for the use subgroup; and

Fn (x) = the cumulative frequency distribution of the original predictor at value *x* for the non-use subgroup.

The next step is to reformulate the 6 predictor variables into 6 dummy variables (1,0), based on the cutoffs of the original predictor to maximize the power of discrimination[1]. Then, the 6 predictor dummies (woman's age, education, number of living children, work status, urban-rural residence and husband's education) are converted into ratio form based on the cutoff points in order to reduce both the variability and skewness of the distributions, expressed in terms of proportions ($P_1, P_2, \ldots P_6$). Similarly, the values of the contraceptive use are converted into the prevalence rate, P_{cc}, at cluster level. This is done from cluster 1 to m (=301) as shown below.

Cluster No.	P_1	P_2	P_3	...	...	P_{6c}	P_{cc}
1	P_{11}	P_{21}	P_{31}	...	...	P_{61}	P_{c1}
2	P_{12}	P_{22}	P_{32}	...	...	P_{62}	P_{c2}
...	...	...	...	...	...	...	
...	...	...	...	...	...	...	
m	P_{1m}	P_{2m}	P_{3m}	...	...	P_{6m}	P_{cm}

Using multiple regression, the functional model is fitted with the contraceptive prevalence rate, P_{cc} and the 6 proportions ($P_1, P_2, \ldots P_6$), with the cluster as the unit of analysis. That is we fit,

$$P_{cc} = \beta_0 + \beta_1 P_{1c} + \beta_2 P_{2c} + \ldots + \beta_6 P_{6c} + e_c$$

where,

P_{cc} = the observed prevalence rate in cluster c (= 1, 2, . . . , m);
β_0 = the intercept;
β_1 = the coefficient for predictor P_{1c};
⋮
β_6 = the coefficient for predictor P_{6c}; and
e_c = the error term for cluster c.

Using the Statistical Package for Social Sciences (SPSS), the regression coefficients β_0, $\beta_1, \ldots \beta_6$ are obtained. The coefficients are then used with the previous proportions to predict the prevalence rates at cluster levels.

The predicted values are therefore given by,

$$\hat{P}_{pcc} = \hat{\beta}_0 + \hat{\beta}_1 P_{1c} + \beta_2 P_{2c} + \ldots + \hat{\beta}_6 P_{6c}$$

These predicted values are then translated into district level estimates, using the following formula:

$$P = \frac{\Sigma b_c w_c P_{pcc}}{\Sigma b_c w_c}$$

[1]Discriminant analysis is a multivariate statistical analysis that uses linear combinations of the independent variables (predictors) to classify cases into groups. It identifies that level of each predictor which explain the differences between user and non-user of contraceptive.

where,

P = the prevalence rate at the district level;

b_c = the number of women in cluster c;

W_c = the sampling weight for cluster c;

P_{pc} = the predicted prevalence rate for cluster c; and the sum is taken over all clusters in the given district.

The district level estimate of contraceptive prevalence rates obtained through regression method is presented in Table 1. Note that in each division there are some districts which have dramatically high or low rate of CPR. The regression method smoothed out (brought in line with divisional variation) these extreme values. Compared to direct estimate, the regression method slightly reduced the inter-district variability. The estimates of the relative errors in the regression method are, in general, lower than those in the direct method. In this procedure, estimates of greater than 86 percent have rather smaller relative errors than that of direct estimates. That is, Bhola in Barisal division, Maulvibazar in Chittagong division, Rajbari, Netrakona and Sherpur in Dhaka division and Khulna, Bagerhat in Khulna division have large sampling errors as compared with direct estimates. All the districts in Rajshahi division have smaller relative errors than that of direct estimates. The accuracy of the estimates appears to have improved by regression method and the estimates obtained by this method seem to be reasonable.

Conclusion

All the methods considered in this study provide consistent estimates (the larger the number of PSUs in the area, the better the estimate). Of the three methods considered to obtain the CPR, the regression method was found most suitable. It overcomes almost all the difficulties of other methods of small area estimation. Most of the regression estimates have smaller relative errors than that of direct estimates. Unlike the synthetic estimation procedure, no auxiliary data are needed in regression approach and the assumption of homogeneity is not required. In this approach, more than one auxiliary variable i.e. all the variables related to the study variable can be used at a time. This has introduced one more advantage to regression method. Another advantage of regression method is that in this method there is no scope of biasness. The regression method allows for estimates of sampling errors while the standard synthetic method do not. So, finally we can say, of the three methods used to estimate contraceptive prevalence rates for districts (small area), the regression method is found most suitable.

References

Goldstein, M. and Dillon, N., 1978, *Discrete Discriminant Analysis.* New York: John Wiely, pp 70-71.

Gonzalez. M. E.. 1973. Use and evaluation of synthetic estimates. *Proceedings of the American Statistical Association (Social Statistics Section)*, December 27-30, pp. 33-36. Washington D.C.

Mitra, S. N., Ali, M. N., Islam, S., Cross, A. R. and Sana, T., 1994, *Bangladesh Demographic and Health Survey 1993-94*. Mitra and Associates, Dhaka, Bangladesh.

Muhuri, P. K. and Rutstein, S. 0., 1994, *Socioeconomic, Demographic, and Health Indicators/or Sub-nalional Areas.* DHS Comparative Studies No. 9. Calverton, Maryland : Macro International Inc.

Platek. R.. Rao, J. N. K.. Samdal, C. E., and Singh, M. P., 1987, *Small Area Statistics.* Wiely. New York