How the sample size was arrived at and why it is adequate

The sample size for the purpose of the study was determined by taking into consideration a number of qualitative and quantitative factors. The calculation of an appropriate sample size relies on a subjective choice of certain factors and sometimes estimates. The important qualitative factors that were taken into consideration were the nature of the research, the nature and number of variables, the nature of the analysis, sample sizes used in other studies, resource constraints and attrition of respondents. Since the study involved descriptive surveys taking a large number of variables, it was necessary to take a large sample. Moreover, it was considered if the sampling size would be large, the sampling variation or standard error would be small. However, the cost of collection of data for large sample and the accessibility to the groups and availability of group leaders were also considered during the pilot survey. After receiving assurances from the CDS and municipality officials that they would help in identifying and accessing the groups, it was decided to go in for a large sample.

On the quantitative side the size of the sample was based on the confidence interval approach. For this purpose the level of precision, i.e., the maximum possible difference D between the sample mean of group income and population mean was taken as +/- Rs 5. The level of confidence was specified at 95% for which the associated z value was 1.96. The standard deviation of the income of groups of the population was estimated on the basis of a pilot study [1] and calculated to be 68 and rounded off to 70. The sample size using the formula for standard error using the formula $n = (\sigma^2 z^2) / D^2$ was found out to be 752. Since the sample size exceeded 10% of the population in 2005-06 the finite population corrector (nN/ n+N -1) was used to arrive at a sample size of 452. However, any sample size calculation is based on the total number of subjects who are needed in the final study. In practice, eligible subjects will not always be available for response, especially in case of longitudinal designed studies and it will be necessary to approach more subjects than are needed in

the first instance. In addition, even in the very best designed and conducted studies it is unusual to finish with a dataset in which complete data are available in a usable format for every subject. Subjects may fail or refuse to give valid responses to particular questions, physical measurements may suffer from technical problems, and in studies involving follow up there will always be some degree of attrition. To take care of this problem the incidence rate and completion rate was taken into consideration for determining the initial sample. Incidence rate refers to the rate of occurrence or the percentage of persons eligible to participate in the study In this case since all group leaders/ members of the TCS and DWCUA groups were eligible, the incidence rate was taken to be 100% or 1. The completion rate denotes the percentage of qualified respondents who complete the survey. In this research it was expected that the completion rate of 90% since assurances had been received from CDS and municipal officers. This meant that the initial sample size should be 1.11 times the final sample size. The initial sample size was therefore 452 multiplied by 1.11 which equals 500. For the purpose of identifying the groups, the names and addresses were collected from the respective CDS and municipalities and a random number generated from computer was assigned to each groups. The 500 groups were then chosen randomly from each CDS and municipality. The questionnaires were administered to the group leaders/members usually when they assembled in CDS meetings and also in the localities in which the group members resided. Due to some problems in responses 18 questionnaires had to be rejected and the number of valid responses was initially 482. Since the study involved use of longitudinal design for study of impact a part of the questionnaire had to be administered to the respondent in 2008-09 again. In this survey 24 groups could not be traced and had to be eliminated from the sample. The sample was therefore finally taken as 458 for the study.

---

1. *"The standard deviation of the population might be estimated by conducting a pilot study. Alternatively it might be estimated on the basis of the researchers judgement"* Malhotra N. in *Marketing Research- An Applied Orientation*, *Pearson Education*, *Fourth edition* 2005, pp346-347