6. STATISTICAL METHODS FOR ESTIMATION

The analysis was performed on the data collected among children without BCG scar. The following criteria were used for estimating the prevalence of infection:

- Method I : Tuberculin reactions of size equal to or more than the identified cut-off point, from the frequency distribution of reaction size were considered attributable to infection with tubercle bacilli.
- Method II: Identifying the mode of reaction sizes attributable to tuberculous infection and estimating prevalence by mirror image technique.

The statistical procedure adopted for the estimation of the prevalence of infection using the above methods has been uniform for all the zones. The data interpretations as well as the methodology of estimation were strictly as per the statistical design adopted in the survey.

District Estimate

The district level estimates of the prevalence of infection (P_d) separately for rural and urban areas were estimated as the weighted average of cluster proportions, the weights being the inverse of probability of selection of the cluster, since the selection of the sample cluster within the district was using PPS sampling. The formula for estimation of the district prevalence, thus was:

$$P_d = \frac{\sum_{i=1}^{n} \frac{P_i}{\pi_i}}{\sum_{i=1}^{n} \frac{1}{\pi_i}}$$

Where P_i is the proportion of children infected in the ith cluster and π_i is the initial probability of selection of the cluster i.e. ratio of the population of ith cluster to the district population and n the number of clusters.

Zonal estimate (by stratum)

The zonal level estimates of the prevalence of infection for the rural and urban strata were estimated by pooling the respective district estimates, using the proportion of district population in the zone as the weight, since the sample size varied between districts. The formula for estimation can be expressed as:-

$$p_s = \frac{\sum_{i=1}^k w_i p_{di}}{\sum_{i=1}^k w_i}$$

Where p_s is the zonal level estimate of prevalence of infection for the respective stratum, p_{di} is the proportion of infected children in the ith district, w_i is the corresponding weight i.e. the proportion of respective population of the district to the total population of the zone and k, the number of districts.

The standard error of the stratum estimate was calculated as:-

$$S = \sqrt{\frac{\sum_{i=1}^{k} W_{i}^{2} (P_{di} - P_{s})^{2}}{(\sum_{i=1}^{k} W_{i})^{2}}}$$

Zonal estimate

The prevalence of infection for the zone (P_z) was estimated similarly by pooling the estimates for rural and urban strata, the weights being the proportion of rural and urban population in the zone. The standard error (S_z) was calculated using the following expression :-

$$S_{Z} = \sqrt{\frac{\sum_{i=1}^{2} v_{i}^{2} s_{i}^{2}}{\left(\sum_{i=1}^{2} v_{i}\right)^{2}}}$$

where s_i is the standard error for rural or urban strata and v_i is the proportion of population in the respective stratum.

ARTI was computed from the estimated prevalence of infection (P) by using the following equation: $ARTI = 1-(1-P)^{1/A}$, Where A is the mean age of the children test read. It was calculated after adding 0.5 years to the age of each child in completed years.

Chi-square (χ 2) test with continuity correction was used to test the significance of differences between proportions and p-values of <0.05 were considered significant.

The data was analyzed using SPSS software.