Introduction

Reliable information on the epidemiological situation of a disease is vital to the development of appropriate control measures and evaluation of the impact of intervention programmes. Tuberculosis disease surveys are impracticable, as a very large population is required to be investigated. In view of low case-finding efficiency of District Tuberculosis Programme (DTP) and the fact that a significant proportion of tuberculosis patients get diagnosed by the private sector, the current system of monitoring of DTP in the country does not provide information on the epidemiological situation in a community.

Tuberculosis situation in any area is related to the rate of transmission of tubercle bacilli from infectious cases to the susceptible population. The probability of getting newly infected during the course of one year is defined as the average Annual Risk of Infection (ARI). It can be derived from the observed prevalence rate of infection obtained by tuberculin testing the children below 10 years of age without BCG scar. Estimation of ARI requires a much smaller population to be investigated and is thus feasible and affordable. To obtain a reliable estimate of ARI, selection of a representative sample of children with adequate sample size is one of the primary pre-requisites. This paper discusses various factors involved in estimation of sample size for tuberculin surveys.

Factors to be Considered for Sample Size Estimation

a) Expected prevalence rate of infection in the study population
b) Desired precision
c) Thesudy population, area (Rural/Urban), cluster size

Expected prevalence rate of infection

Prevalence of infection in the study age group (0-9) years observed in an earlier survey carried out in the same or adjacent area could be considered as the expected prevalence rate (p). When such information from the study area is not available, the value of p may be obtained from tuberculin surveys carried out in other characteristically similar areas of the country.

In some of the tuberculin surveys carried out in different parts of the country, the reported prevalence rate of infection for the age group 0-9 years varied between 5 to 10%. Therefore, for an area where information on prevalence rate from earlier surveys is not available, it is desirable to assume a lower expected prevalence rate of 5% which increases the statistical validity of the sample size.

Precision

The precision denotes the difference of value obtained in the study sample from the true prevalence rate of infection in the population i.e., true prevalence in the population-observed prevalence in the study sample. The larger the sample size, more is the precision. However, in determining the sample size, it is required to strike a balance between precision and operational feasibility including the cost. Depending on the resources available for conducting the study, the level of precision required is decided before estimating the sample size. It is reasonable to obtain an estimate of prevalence of infection in the study sample within 10 percent (relative precision) of the true prevalence in the study population. This means that when true prevalence rate of the population is 10%, the acceptable prevalence rate obtained in the study group could be between 9 and 11 percent. Similarly when true prevalence is 5%, it will be between 4.5 to 5.5 percent. When constraint of resources exist, the level of precision may be reduced upto 15 percent.

Study population, area and cluster size

The choice of study population for an ARI study would be BCG unvaccinated children below 10 years of age, in whom the prevalence of mycobacteria other than tuberculosis (MOTT) is minimum. When a house to house survey is...
planned, children of 0-9 years of age comprise the study group. However, when the study is intended to be carried out in schools, the desired age group could be 6 to 9 years.

When an ARI study is planned for a rural area, a widespread coverage of the study area enhances the validity of the data collected. In doing so, factors of operational convenience need to be kept in mind. Therefore, before undertaking the survey, the ability of a field team to register and examine the number of eligible children per day in a village context by house visits is estimated. Based on this, a village with a coverable study population is considered as a cluster (the number of children who could be covered in a village per day is called as cluster size). Depending on the estimated sample size required for the survey and the cluster size, the number of clusters to be covered for the study is calculated.

The list of villages located in a study area could be obtained from the recent census book brought out by Government of India. From this list, the number of clusters/villages required are selected by a simple random sampling technique.

On verification, if school attendance of study group in rural area is found to be above 70-75%, a school based survey could be carried out, which is operationally convenient and cost effective when considered against house to house survey. As suggested earlier, when school survey is planned, the study population should be 6 to 9 years aged. BCG unvaccinated children and the cluster size under school survey would be the number of eligible children who could be covered per day in each school. On obtaining the list of schools located in the study area from the education department, the number of clusters/schools required for the study is selected as explained above.

It is desirable to cover equal number of children in each cluster. On practical considerations if all children in a cluster are required to be covered, then a weighted analysis has to be carried out for unequal size of population covered in each cluster.

Since carrying out house to house survey in urban area is difficult, a school based survey is preferable.

**Design factor**

The basic sampling unit for a tuberculin survey is a child. As the information of all children in the study area is not available, it is required to opt for cluster sampling. In order to obtain estimates of ARI as precise as with simple random sampling of children, the sample size for cluster surveys is required to be increased by the appropriate design factor (D), which is to compensate variations occurring within the clusters and between the cluster. Design factor may be estimated from the data from tuberculin survey of similar design carried out previously, as the ratio of the cluster sample variance to the variance as if it were a simple random sample. If such a data is not available, D has to be approximated. It is appropriate to have a design/factor of 2 to 2.5 for tuberculin surveys. On completion of the study, value of D can be worked out from the data obtained.

**Estimation of Sample Size**

Based on the available information, sample size required for an ARI study is calculated using the formula:

\[ n = \frac{Z^2 \cdot (1 - \pi) / \pi^2}{\pi(1 - \pi)} \]

Where \( n \) is sample size; \( \pi \) is expected prevalence of infection; \( \alpha \) is the level of significance, \( Z \) represents the standard errors within which most of the sample proportions shall fall, assuming normal distribution. If it is chosen as 1.96, 95% of all sample proportions are expected to fall within 1.96 times the standard error of population proportion.

With expected prevalence of 10%, a sample size of 3452 would be required for a random sample of the children to be investigated in order to obtain an estimate of prevalence of infection within 10 percent of the true value with 95% confidence. It has to be modified for cluster sampling by multiplying with the value of design factor. Taking a design factor of 2, about 6915 children would be required to be investigated (Table).

Depending upon the expected prevalence of infection in the study age group and required precision, the appropriate sample size may be obtained from the following table.

**Table Sample size* for tuberculin surveys**

<table>
<thead>
<tr>
<th>Expected prevalence rate of infection (p)</th>
<th>Sample size for ( \pi = 10% )</th>
<th>Sample size for ( \pi = 12.5% )</th>
<th>Sample size for ( \pi = 15% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>14,598</td>
<td>9,343</td>
<td>6,488</td>
</tr>
<tr>
<td>6%</td>
<td>12,037</td>
<td>7,704</td>
<td>5,350</td>
</tr>
<tr>
<td>7%</td>
<td>10,208</td>
<td>6,533</td>
<td>4,537</td>
</tr>
<tr>
<td>8%</td>
<td>8,836</td>
<td>5,655</td>
<td>3,927</td>
</tr>
<tr>
<td>9%</td>
<td>7,769</td>
<td>4,972</td>
<td>3,453</td>
</tr>
<tr>
<td>10%</td>
<td>6,915</td>
<td>4,426</td>
<td>3,073</td>
</tr>
</tbody>
</table>

*considering a design factor of 2
Sample Size for Repeat Surveys

Once reliable baseline estimate of tuberculous infection is available for an area, repeat surveys provide information on the trend of disease in the community. A significant increase in ARI indicates introduction of new risk factors for transmission of infection and requires further detailed epidemiological investigations in order to identify the factors responsible and reformulate the control strategy, if required. On the other hand, decline in ARI reflects the impact of control interventions. However, an inefficient case-finding and treatment programme is unlikely to bring a significant decline in ARI. Hence a very large sample size would be required to detect a small change in the disease situation, say after a period of 5 years. The current tuberculosis situation in the country is akin to this.

When comprehensive tuberculosis programme has been carried out in an area, a significant decrease in ARI rate upto 20% has been reported. Even in developing countries, with improvised control measures, a decline in risk of infection of about 5 percent has been predicted. In such situations, repeat ARI surveys involving about 3000-5000 children, after an interval of 5 years with the observed baseline prevalence rates of 5 to 10 percent, can measure the impact of interventions.

Conclusion

The epidemiological trend of tuberculosis in a community could be assessed by carrying out tuberculin surveys in an affordable size of population, in contrast to a large population of approximately one lakh required to estimate the prevalence of bacillary tuberculosis disease in a community and still larger population required for following its trend.

Acknowledgments

We are thankful to Mrs.Vijayalakshmi for secretarial assistance.

References


