

## PLANNING AND CONDUCT OF EPIDEMIOLOGICAL SURVEYS\*

G. S. ACHARYULU

(From Madanapalle Tuberculosis Research Unit, Madanapalle)

Pulmonary Tuberculosis is the most common form among the different forms of tuberculosis. The main indices useful in the epidemiological study of this disease are the prevalence and incidence rates of (i) persons infected with tubercle bacilli, (ii) persons excreting tubercle bacilli and (iii) suspect cases of tuberculosis (diagnosis not confirmed bacteriologically). With the occurrence of primary drug-resistant cases, knowledge of the prevalence and incidence of such cases has attained considerable importance.

There are two types of epidemiological surveys, prevalence surveys and longitudinal surveys. In prevalence surveys, the observations refer to a specific point in time. In longitudinal surveys, the observations are repeated at different points in time. Both prevalence and incidence rates can be estimated from these surveys. They are useful for making future projections of the tuberculosis problem and in evaluating the effects of different tuberculosis control programme.

Longitudinal surveys can be classified as experimental surveys and non-experimental surveys. The objective in experimental survey is to assess the value of or to compare the effects of different T.B. control programmes. In non-experimental surveys, no such assessments or comparisons are contemplated.

### Considerations in Planning the Surveys

The first step in planning an epidemiological survey is to define the objectives of the survey. This involves the specification of the purpose of the survey and the epidemiological indices which are to be estimated, and how the survey results are going to be made use of. On the basis of these indices it will be determined whether the survey should be a prevalence or a longitudinal type.

In experimental survey, the selection of the individuals into the study depends upon the type of population to which the experimental results are to be later applied. There are two situations : the first is, the population is supposed or known to possess certain characteristics which will interact, favourably or unfavourably, with the T.B. control measures, which are to be compared in the experiments ; the second is when there is no knowledge about any such

characteristics. In the latter situation, a sample reasonably representative of the population is selected according to practical convenience for undertaking the survey. In the former situation, the population will have to be divided into a number of strata and a representative sample will have to be selected from each stratum. An example of such a situation is provided by the presence or absence or non-specific sensitivity in the population in a study designed to investigate the protective effect of BCG Vaccination, as there is reason to believe that non-specific sensitivity offers some protection against tuberculosis.

Special problems which are to be considered in planning experimental surveys are :

- (i) incorporation of suitable 'controls',
- (ii) method of randomization and
- (iii) determination of the size of the experimental population.

In situations, when there is doubt regarding the effects of T.B. control measures or when the interest is in evaluating their effects, 'controls' must always be introduced in the experiment. If, on the other hand, there is no doubt regarding the effects of control measures and the interest is only in comparing the effects of different control measures, there is no necessity for 'controls'. If valid comparisons are to be made, the experimental units (an experimental unit consists of a single individual or any well defined group of individuals) should be randomly allocated to the different programmes. There are many methods of randomization and they make use of the knowledge, if available, of the inherent variability of the experimental material which is relevant to the investigation. The purpose of these methods is to increase the efficiency of the experiments by reducing the experimental error. The factors to be considered in the choice of a proper method of randomization are :

- (i) the types of control programmes to be compared,
- (ii) available knowledge of the inherent variability of the experimental material and
- (iii) practical considerations in the conduct of the survey such as actual process of randomization etc.

\* Paper read at the 24th National Conference on TB and Chest Diseases held in Trivandrum in January 1969.

Size of the experimental population will be determined by making use of the estimates of of experimental errors obtained from previous surveys.

There are two methods of undertaking prevalence surveys or non-experimental surveys ; complete survey method and sample survey method. The objective in both the methods will be to obtain knowledge of the distributions, aggregate or mean values of one or more characteristics of a well defined population. In complete survey method, observations are made for every individual in the total population, whereas, in sample survey method, observations are confined to a certain number of individuals in the population. On the basis of the selection of the individuals, samples can be divided into random and judgement samples. In random samples the individuals are selected according to the laws of probability while, in judgement samples, selections are not done according to these laws. The National Tuberculosis Sample Survey, conducted during 1955-'58, under the auspices of the Indian Council of Medical Research, is an example of random sample survey. The European Tuberculin Survey undertaken during 1965-'66 under the auspices of the International Union Against Tuberculosis and the First Drug-resistance Survey (1964-'65) conducted under the auspices of the Indian Council of Medical Research are two examples wherein judgement samples have been made use of.

The random sampling method makes it possible not only to estimate the population characteristics but also to estimate the errors involved in those estimates. This method will ensure that the samples will be representative of the population. On the contrary, there is no guarantee that the judgement samples will be representative of the population. Further, the errors involved in the estimates of the population characteristics cannot be estimated from judgement samples. Another disadvantage is, that it will not be possible to make valid comparisons of the estimates obtained from one judgement sample with those obtained from another. In view of these considerations, random sample surveys are always to be preferred. Judgement samples should be made use of only if it is not possible, owing to practical difficulties, to undertake random sample surveys.

With regard to the choice between random sample surveys (hereafter referred to as sample survey) and complete surveys, sample surveys are always to be preferred as they have several advantages over complete surveys. These are :

- (i) reduction in the cost of the survey,
- (ii) the speed with which the estimates can be obtained and
- (iii) a greater accuracy in the results.

Greater accuracy will be obtained as it will be possible to control the various sources of errors in collecting the data by employing technically trained and experienced personnel, when the survey is conducted on a limited scale.

The undertaking of sample survey requires:

- (i) the construction of a sampling frame,
- (ii) the choice of a sampling design and
- (iii) the fixation of the sample size.

The sampling frame consists of a list of sampling units in the total population without omission and/or duplication and this is essential for drawing random samples. The sampling unit may consist of an individual or all members in the household or a group of households in a contiguous area or any other well defined group. The sampling design stipulates the method of drawing the sampling units into random sample. For the same sample size, certain sampling designs will be more efficient than others, efficient in the sense of containing comparatively less sampling error. Although it is possible to minimize the sampling errors by choosing efficient sampling designs and suitably increasing the sample size, in practice, the decision regarding the sampling design will have to be taken on the basis of the sampling frame that is available and mainly in view of the operational convenience of the field work. The size of the sample will be fixed by the consideration of the cost of the survey and the margin of error allowed in the estimates.

### **Types of Data collected**

Specific data to be collected in any particular survey will be decided by the specification of the layout of the final statistical tables required from the survey. The types of data that are generally collected in epidemiological surveys relate to (i) census, (ii) tuberculin tests, (iii) X-ray examinations, (iv) bacteriological examinations, and (v) anti-tuberculosis measures. A discussion of the considerations in collecting some of these data may be relevant here.

### **Tuberculin Tests**

Although the importance of this test in measuring the prevalence and incidence of infection is well known, its value is restricted

by the use of BCG Vaccination. Hence, this factor, that is, the extent to which the population concerned was covered by BCG Vaccination is to be considered before deciding to use this test for measuring the epidemiological indices.

### **X-ray examinations**

In countries like India, children in the younger age groups constitute a considerable proportion of the total population. In view of the low prevalence and incidence of the disease among children, it is sometimes desirable to exclude them from these examinations. Such a procedure will reduce the work-load considerably and at the same time the resulting loss in precision of the estimates will be negligible. Regarding the interpretation of X-ray films, a suitable code to meet the specific requirements of the survey will have to be devised, as no internationally accepted code\* has so far been evolved for this purpose. Two independent readers are considered essential.

### **Considerations in the conduct of the Survey**

For the success of the survey, it is essential to have well trained and experienced personnel. When such personnel are not available, recruitment and training programmes will have to be undertaken well in advance. The various cards and forms to be used for collecting the data should be designed. A detailed protocol for the conduct of the survey should be prepared. This protocol should:

- (i) Specify the purposes and objectives of the survey;
- (ii) define the populations to be surveyed;
- (iii) describe the various methods of examinations to be conducted along with the criteria of classifications to be adopted;
- (iv) define the concepts to be used such as case of tuberculosis etc.;
- (v) contain detailed instructions to the staff working in different sections such

Efforts are being made by a specially appointed international committee (The Ad Hoc Committee for the study of Classifications and Terminology in Tuberculosis under the auspices of IUAT in cooperation with WHO), to evolve a code for interpretation which will be acceptable internationally and which will minimize the errors of interpretation.

as census taking, tuberculin testing etc.

The instructions to the field staff should be unambiguous and should contain typical illustrations. If necessary, a pilot survey may be undertaken. Such a survey will be useful (i) in the development of the field procedures, (ii) for testing the suitability of the various cards and forms designed and (iii) in giving training to the field staff.

An important aspect, particularly in the conduct of experimental surveys, must be mentioned here. This is to ensure that groups of individuals allocated to different control measures to be compared are treated alike throughout the survey with regard to the admission of the individuals into the groups, intensity of the diagnostic examinations and other factors except the actual control measures. How any such factor can introduce bias into the study may be illustrated by an example. In a BCG Trial, if places injections are not given to 'controls', it will not be possible to identify the group among the controls corresponding to the group of persons who refused BCG Vaccination. Another important aspect is that the identity of the control measures given to an individual should not be revealed either to him or to those who will have to assess the effects of control measures.

### **Sources of Errors and Measures of Control:**

In order to get valid and reliable estimates it is essential to minimize the errors which arise from several sources. These are broadly divided into sampling and non-sampling errors. Sampling errors occur only in sample surveys, whereas non-sampling errors occur both in sample and complete surveys. While the sampling error, as mentioned earlier, can be kept within predetermined limits by suitable measures, it will not generally be possible to identify all the sources of non-sampling errors and to estimate the extent to which they affect the final estimates.

Non-sampling errors arise on account of several factors and some of them are discussed below :

1. *Failure to examine the complete study population* : This is called the error due to non-response. One method suggested to obtain unbiased estimates of the population characteristics is to examine a random sample of persons who failed to report for the examinations earlier. This procedure will not be practicable in many situations. The other way of reducing this error is to minimize the percen-

tage of non-response by intensive efforts and to note down the reasons for non-response, so that an analysis can be made later to give an idea as to how the estimates are going to be affected.

2. *Biases* : There are two types of biases, systematic and unsystematic. The occasional failure to detect every bacillary case with single spot sputum specimen is an example of systematic bias. Such a procedure will lead to consistent under-estimation of the number of bacillary cases. The extent of under-estimation can be ascertained by conducting special studies or from the results of such studies made elsewhere. If the resources permit, this error can be reduced by better methods of sputum collection.

The effect of unsystematic bias is sometimes over-estimation and some times under-estimation. The unconscious preference for round numbers, generally observed in reading the tuberculin reactions and in reporting the ages, may be given as examples of unsystematic biases. The effects of these errors can be reduced to some extent, by suitably grouping the observations in preparing the frequency distribution when the data are tabulated.

3. *Errors in classification* : Generally, tuberculin positive reactors who are supposed to have been infected with tubercle bacilli are denoted to be those whose reactions exceed a certain size, for example 6 mm. or 8 mm. The dividing point between negative and positive reactors varies from one study to another and maybe anywhere between 6 to 15 mm. This type of definition is known to involve two types of errors, viz., inclusion of persons who have been actually infected with tubercle bacilli and exclusion of persons who have been actually infected with tubercle bacilli. The magnitude of these two types of errors depends upon the magnitude of the prevalence of non-specific sensitivity in the community. As long as the same definition is

used for this purpose, provided that other conditions are the same, comparison of estimates from one group with another will be valid. But, when the interest is in estimating the absolute value of the magnitude of infection with tubercle bacilli, efforts should be made to evolve an objective criterion for this purpose.

4. *Errors in data processing and tabulation*: These errors can be reduced by giving clear instructions, by employing well trained staff and by exercising constant supervision over them.

To sum up, the first step in planning a survey is to provide a clear and unambiguous statement of its purposes and objectives. Planning requires consideration of all important aspects of the survey, many of which are inter-dependent. The application of statistical methods is essential to get valid and reliable estimates and for making valid comparisons. Recognition of the types of errors which might occur in survey will not only lead to better planning and execution of the survey, but also to better understanding of the reliability of the estimates obtained from the survey.

#### REFERENCES

1. WHO Expert Committee on Tuberculosis, 8th Report, 1964 WHO Technical Report Series, No. 290.
2. Protocol of the Coordinated European Tuberculin Survey, *Bull. WAT*, 1965, No. 1, Vol. 36, 42-44.
3. Report of the Committee on Epidemiology and Statistics, *Bull. WAT*, 1966, No. 1, Vol. 38, 71-74.
4. Gangadharam, P.R.J., Proceedings of the 21st Tuberculosis & Chest Diseases Workers' Conference, 1966, 104-112.
5. Report of the Ad Hoc Committee for the study of Classification and Terminology in Tuberculosis, *Bull. WAT*, 1965, No. 1 Vol. 36, 55-72.