

APPENDIX 5

TYPES OF SAMPLING PROCEDURES

Sampling is a procedure for choosing selected units from a larger population and drawing conclusions about the larger population from them. "Population" is a statistical term used to describe the total number of units from which to sample; these may be individuals in a village, mosquitoes in a house, health centres in a district, days in the year etc. There are certain rules to follow. Four approaches which satisfy these rules are described below. In each case the basis on which the sampling was made should be explicit. If units are excluded from the original population, this should be noted and the reasons given. A statistician may need to be called in to assist in selecting a sampling strategy and sample sizes, and in analysing the data.

Simple random sampling This is the preferred technique when all the units in a population can be feasibly listed. It can be used, for example, to select houses in a village in order to study mosquito abundance.

- obtain a list (arranged in any order) of all the units from which the sample is to be selected (i.e. from the population) and number them;
- decide on the size of the sample (see below);
- select the required number of units at random using a table of random numbers.

Systematic sampling If numbering all the units is difficult, the following technique is somewhat simpler:

- obtain a list of all the units from which the sample is to be selected (say n units in the population);
- decide on the size of the sample, say s units in the sample (see below);
- calculate the ratio $s/n (= k)$. Select every k th item on the list, starting at any point. For example, say a sample of 50 patients was wanted (i.e. $s = 50$) out of 2000 attending a clinic in a year (i.e. $n = 2000$). Then $k = 50/2000 = 1/40$, and every 40th patient would be selected.

Cluster sampling If the costs of all the insecticide spraying units in the country are of interest, studying a random (or systematic) sample would probably require a lot of travelling and effort since these units are likely to be widely dispersed. An alternative approach is first to select a random sample of districts and then look at the spraying units (or a random sample of them) in those districts. This is called cluster sampling (the units are clustered in a limited number of districts). It gives less precise results than random sampling but can have major logistic advantages.

Stratified sampling It may be desirable to ensure that the sample includes certain kinds of units which, because there are relatively few of them, may not be selected in a random sample (for example, vector control units operating in the coastal and mountainous regions). To do this, the total population (i.e. all vector control units) is first divided into subgroups (in this case coastal and mountain) and then random or systematic sampling done in each subgroup.

Sample size Determining appropriate sample sizes is not always easy. The larger the sample size the more accurate the measurement (assuming that a representative population was chosen originally). At the same time, however, the larger the sample size the more expensive and time-consuming it is to collect the data. The greater the variation that exists between the units, the larger the sample size needs to be to achieve reasonable precision. Where there is little variation (for example, perhaps, in the costs of running a standardized vector control unit), smaller sample sizes can be used. Consult a statistician for advice on sample sizes.

