Chapter 2

METHODS OF SAMPLE SELECTION AND COLLECTION

B. DUFOUR, F. MOUTOU & A. RODOLAKIS

2.1 SUMMARY

In presenting methods of sample selection and sample collection, the words are defined in absolute, relative precision and accuracy. Random samples are differentiated from empirical samples and some simple rules of sampling are proposed and explained. The main types of samples to be collected are described together with the precautions that should be taken during their handling and transportation.

2.2 INTRODUCTION

During the study of the epidemiology of animal diseases, comprehensive prophylactic campaigns or investigations may be undertaken involving large numbers of livestock in one region or even one country, with one or more species being sampled. Under such conditions, the results obtained from analysis of the samples can, with some allowances, be directly exploited and interpreted particularly if the samples have been collected under the best possible conditions.

Unfortunately this situation does not always occur. For reasons of cost and lack of personnel and materials, sampling is often restricted to only one part of the population, which represents a sub-sample of the whole. It is then necessary to question how, under what conditions and with what precautions, the results obtained from analysis of these samples represent the original population.

In the first instance it is necessary to know if a subsample in fact represents the original population. To answer this question a certain number of rules must be followed and these are presented in this chapter, together with recommendations for carrying out good sampling techniques. However, it is first necessary

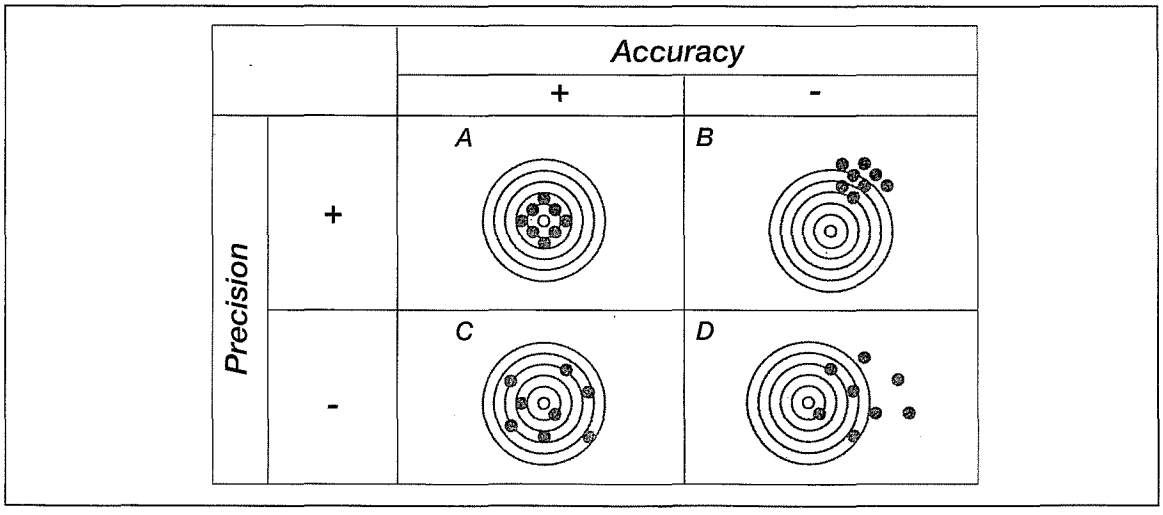

Figure 2.1 : Accuracy and precision

to define some of the terms commonly used in this context.

2.3 METHODS OF SAMPLE SELECTION

2.3.1 Sample quality

2.3.1.1 Definitions

Two important terms associated with sample selection are accuracy and precision. Accuracy is the quality of the agreement between the measured value and the true value, whilst precision is the measure of deviation of the values of a repeated measurement from the mean value. These two parameters are illustrated in Figure 2.1. Both can change independently of the other which results in four possible scenarios. The situation in diagram A corresponds to the desired accurate and precise result. In diagram B, the result is still precise but it has become inaccurate. This type of result could be obtained from a measuring device on which the zero has been poorly calibrated. In the situation in diagram C, precision has not been obtained but the accuracy, on average, is good. The variation around the mean of the tested value is important, but the true value is effectively included in this large spread. Finally, the last situation in the figure, diagram D, is the one to be avoided at all costs. The result is imprecise and inaccurate and allows absolutely no conclusion to be drawn.

The most common problem is to identify where the risk is located since this cannot always be known in advance. There are two ways to tackle these two risks. To obtain good accuracy, one must select the individuals - animals or flocks - to make up the sample randomly. When this process is effective the pitfall of inaccuracy can be avoided. There is actually little chance that all the individuals would possess the same difference from the mean. To increase the precision the size of the sample should be enlarged so that the range within which the real test value is located will be reduced.

Experience in the field has shown that the selection of individuals to be sampled poses a number of serious problems. A complete list of the flocks, farms, their addresses and sometimes additional assurance that the owners of the animals agree to participate in the investigation needs to be prepared. As the number of samples increases, financial constraints or the capacity of the analytical laboratory to handle the samples will quickly become obvious and impose limitations.

2.3.1.2 Measure of accuracy

If the local agricultural statistics are available, the selection of animals can be undertaken automatically using tables of random numbers or any similar method. Random number tables contain series of totally independent numbers which can be used to select the livestock or the animals to be sampled from the population, after the latter have been given a unique numerical identifier. A number in the table can then be linked directly with the number of an individual to be sampled.

Computers can also be used to generate random numbers and to select animals to be sampled. For example, if one wishes to choose 20 livestock from a total of 100, one numbers the livestock from 1 to 100 and asks the computer to choose 20 numbers between 1 and 100.

2.3.1.3 Absolute precision, relative precision

The absolute precision of a test corresponds to the size of the confidence interval, i.e. to the quantity that is added or subtracted from the mean to obtain the limits of the confidence interval. The relative precision is equal to the absolute precision divided by the value of the estimated proportion.

In a population of 500 individuals where 10%, that is fifty, are infected, selection of a sample of ten individuals will not always give nine healthy individuals and one infected one. Of the ten individuals chosen none may be infected or in another case two, or very rarely three or more may be infected. The estimated percentage of infected animals in the original population will thus be estimated with a certain margin of error. Different selections of samples of ten individuals will each result in a different estimation of the disease prevalence; each of these estimations can be calculated, along with a certain confidence interval in which the true value will be found. The confidence interval can be adjusted to ensure that the probability of the true value being contained within it is equal to any given percentage, the most frequently used value being 95%. Other values could be used, dependent on the scientist's minimum tolerable probability of success. In this case, the confidence interval is approximately + two standard deviations each side of the mean. The precision will increase, and hence the confidence interval will decrease in size as the number of individuals included in the sampling increases and when the value of the disease prevalence in the population approaches 50%, moving away from the extreme values (for a fixed sampling size). Tables of percentage confidence intervals exist for a maximal tolerable probability of success, most often 95%. For high levels of the disease prevalence, the confidence interval will increase in absolute value; the absolute precision then diminishes but the relative precision improves. If the confidence interval is two times the value of the standard deviation about the mean given by the samples, the absolute precision corresponds to the standard deviation and the relative precision to the ratio of the standard deviation to the disease prevalence. If the disease prevalence is calculated as 10% with an absolute precision of 2% (that is 10 + 2%), the relative precision must be 2/10, i.e. 20%, which is ten times less precise.

Finally, it is acknowledged that in some cases the sampled population may be made up of animals from different flocks. The standard of sampling may then differ and the sampling objectives will not necessarily be the same. In one case the proportion of infected animals is being sought in order to identify an infected flock, in another case the proportion of infected flocks is being sought to identify an endemic region. An extreme possibility would be to sample livestock herds in a region then to sample animals within certain of these herds.

2.3.1.4 Objectives

Two questions commonly arise in relation to the objectives of sampling:

* How many individuals must be sampled to detect at least one infected animal or, alternatively, how many individuals must be sampled to confirm a population as uninfected (having a rate of infection below a pre-established threshold) with a tolerable probability of success?

* How can one calculate the minimum or necessary number of individuals to confirm the prevalence of the disease looked for in the sampled number, and how does one choose the level of precision and a maximal tolerable probability (of success)?

The first question corresponds to a calculation of probability. To obtain a result, the level of disease prevalence below which the population is considered as uninfected must be set and the required probability of success specified, usually as 95%. If P is the prevalence of the infective agent in the defined sample, P can also be regarded as the probability of selecting an infected animal from the population and therefore 1-P as that of selecting an uninfected animal. In a sample of n individuals, the probability that none of the animals is infected is thus $(1-P)^n$. The probability that at least one animal is infected is $1-(1-P)^n$, which can then be calculated at 95%, or even 99%. If $C = 1-(1-P)^n$, solving this equation for n yields,

$$n = \frac{\log (1-C)}{\log (1-P)}.$$

To resolve this equation the confidence level, C, must be set, usually at 95%, the maximal tolerable probability (of success). The equation assumes that the size of the sample is small in relation to that of the population, which only actually occurs with numbers greater than about 500, and if the disease prevalence is not too low. Very often, the size of the total population necessitates a correction. The are statistical tables wich show the size of the sample to be taken in relation to the size of the flock and of the disease prevalence for a maximal tolerable probability of success of 95%. It should be noted that the table also incorporates the level of the sensitivity and specificity of the analytical test performed. If in a given population of 400 individuals there is a disease prevalence of at least 1%, according to the table 210 individuals must be selected and must be negative to establish that the population is uninfected with only a 5% risk of being wrong.

To answer the second question, that is the calculation of the minimum number of individuals that must be sampled to calculate the prevalence of an infective agent in a population, it is necessary to have some idea of the prevalence, P, being assessed and to set the desired absolute precision, i, and the minimal tolerable probability of success. In the case of a very large population, the size of the sample, n, is :

$$n = \frac{\varepsilon^2 P(1-P)}{i^2}$$

where e represents the reduced deviation corresponding to the maximum probability of success. In the case of a population which is small relative to the sample, the value of n must be corrected. There is statistical table which show the number of individuals to be selected for sampling as a function of the desired relative precision, the size of the population and the estimate of the disease prevalence. The maximal tolerable probability of failure is 5%.

2.3.2 Different types of sampling

A simple definition of a sample is that it is a subset of a population and as a small part of the whole it must represent certain characteristics of the whole. In selecting a sample, the main risks lie in the possible inclusion of bias which will alter the accuracy and precision. Unfortunately experience has shown that it is very difficult to obtain an unbiased sample. It is therefore important to know where bias is most likely to occur. Bias will vary according to the situation and must be taken into account during the interpretation of the results. Some examples of sample selection are as follows :

2.3.2.1 Empirical sample

An empirical sample, also known as a blind sample, must be distinguished from other sample types. With empirical samples there is no system or a predetermined plan for sample selection and this type of sample should be avoided whenever possible. There is no way of knowing what bias is introduced in this sampling method but, because there is no selection process, accuracy will be poor. In a flock the first 10 animals captured would be sampled whilst in a region samples would be collected from the first 10 farms along the road. A random sample is always better.

2.3.2.2 Random sample

In this situation the sample is composed of selected individuals each with a certain probability, known to the selector, that the sample is representative of the population. Many of the alternative methods available combine knowledge of the population and the studied disease but financial constraints will always affect the choice of method applied.

* Simple random samples. Each individual in a population has the same probability of being selected. The population is homogenous with respect to the considered disease and there is no known influence of factors such as age, sex or farming method.

* Stratified random sample. In this case, factors such as age, sex or the type of breeding are known to have an effect. The population is stratified by, for example, age to create sub-populations of 0-1 years, 1-3 years and more than 3 years. A sample is then randomly selected but the number in each age category is represented by the same percentage as exists in the whole population.

* Clustered random sample. The selected units are of groups of individuals, for example, of livestock, even groups of livestock. This type of sample must not be confused with the flock sample where each basic unit is a flock. With a clustered random sample the unit can be part of a flock, the flock or many flocks. The flock sample is a specific case of cluster sample, also known as a group sample.

* Systematic random sample. This is a practical method of obtaining a random sample which utilises the rule of systematic selection. If 10% of a flock must be sampled and all the animals can be caught individually, every tenth animal that passes the gate of the sheepfold or enclosure is selected and sampled.

2.3.2.3 Choice of method

Ideally the selection method chosen will be the one that is easiest to carry out. However, to provide a good sample it is necessary to have a good knowledge of the disease being studied, the farming conditions and the region where the sampling is taking place so that all the foreseeable bias can be identified. It is better to have a modest sample where all the bias is known than to try to obtain a very sophisticated sample where few of the parameters are controlled. In the field, it is also necessary to know who will do the sampling, who will analyse the samples and to remain very pragmatic.

2.4 SAMPLES

To make a valid diagnosis, samples must be collected correctly, at the right time.

Samples must be clearly labelled and transported to the laboratory as quickly as possible, having been cooled and packaged in a waterproof, insulated container

holding sufficient absorbent material to avoid any loss of liquid in the case of accidental damage. It must always be borne in mind that these biological products are potentially hazardous to humans and therefore any possible leakage from the package must be prevented during transportation. Cool boxes should be used to conserve a low temperature.

All samples must be accompanied by documentation detailing the name and address of the sender, the analyses required and all pertinent information about the abortions. This documentation must be placed in a plastic bag attached to the outside of the package.

Packages must be visibly labelled "Pathological Samples, Fragile, Handle With Caution".

If the fresh tissues are not to be analysed immediately on arrival at the laboratory, they must be stored frozen at -70°C. Fixed tissues must be preserved in a fixative for some days or weeks before being treated.

2.4.1 The placenta

The placenta, when available and not too soiled, is the best sample for the isolation of the majority of the abortive agents. It can also be used for detection of organisms by staining of impression smears or histological sections. Since the entire placenta is difficult and hazardous to handle on arrival at the laboratory, it is preferable to sample cotyledons. Where possible those showing visible lesions should be collected since the degree of infection often varies from one cotyledon to another.

For isolation, 5 or 6 cotyledons together with their associated intercotyledonary membranes should be placed in a sterile container and transported to the laboratory. If the cotyledons are soiled they can be washed beforehand with sterile physiological saline. For isolation:

1. of viruses, cotyledons should be placed in viral transport medium (VTM: see Chapter on Border Disease);
2. of *Campylobacter*, cotyledons should be placed in FPB/glycerol medium (see Chapter 8);
3. of leptospires, cotyledons should be placed in 100ml of 1% bovine serum albumen (BSA) diluent (see Chapter 12).

For histology, sections 0.5cm thick should be taken from other cotyledons showing lesions. These should be placed in glass bottles containing fixative (0.85g NaCl dissolved in 90ml water to which is added 10ml of formol) in a ratio of 10 volumes of fixative to 1 volume of tissue.

For bacteriology, smears should be made by application of cotyledons showing lesions to slides.

2.4.2 Vaginal swabs

Vaginal secretions sampled immediately after abortion by swabbing also provide good samples for isolation of abortive organisms. They are not usually as heavily infected as the cotyledons, but they reflect moderate infection of the placenta. Correctly sampled they are more "appropriate" bacteriologically and less hazardous to the handler. Samples should be collected as soon as possible after abortion. Vaginal excretion, often abundant during the first few days, can decrease rapidly or become intermittent making testing more inaccurate if insufficient numbers of samples are collected.

The swab is made up of a metal wire about 15cm in length, covered with cotton wool at one end. A glass cylinder about 10cm in length and 8mm internal diameter surrounds the speculum and avoids contamination of the cotton wool when the swab is introduced into the

vagina. The whole device should be sterilized by autoclaving in a test tube (Figure 2.2). If unavailable, commercial sterile swabs can be used but these are generally a little short and the cotton wool swab is a little small.

The swab should be sent to the laboratory as it is or preferably in an appropriate transport medium according to the organism to be detected (*Chlamydia*, *Coxiella*, viruses, *Campylobacter*, etc.).

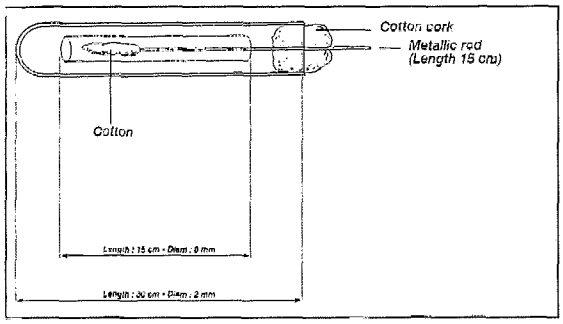

Figure 2.2 : Vaginal swab

2.4.3 Tissues from aborted foetuses and newly dead lambs

The tissues (spleen, liver, kidneys, brain, spinal cord, etc.) must be sampled aseptically as soon as possible after abortion or death. Samples taken must be placed :

- in sterile bottles for isolation or for antigen detection. Samples should be placed in VTM for isolation of viruses, FBP/glycerol for isolation of *Campylobacter* or 1% BSA diluent for isolation of leptospires, as for the placenta ;
- in a glass bottle containing at least 10 volumes of appropriate fixative for histology.

When it is not too autolysed the entire brain should be removed from the foetus (Figure 2.3).

2.4.4 Foetal fluids

The stomach contents, heart blood and peritoneal and pleural fluids can be sampled if the foetus is not too autolysed. These fluids should be sampled aseptically as quickly as possible after abortion using a sterile syringe.

2.4.5 Milk

Colostrum and milk from the two quarters should be sampled aseptically (disinfect the teats and discard the first two jets of milk) for isolation of the abortive agent or antibody detection.

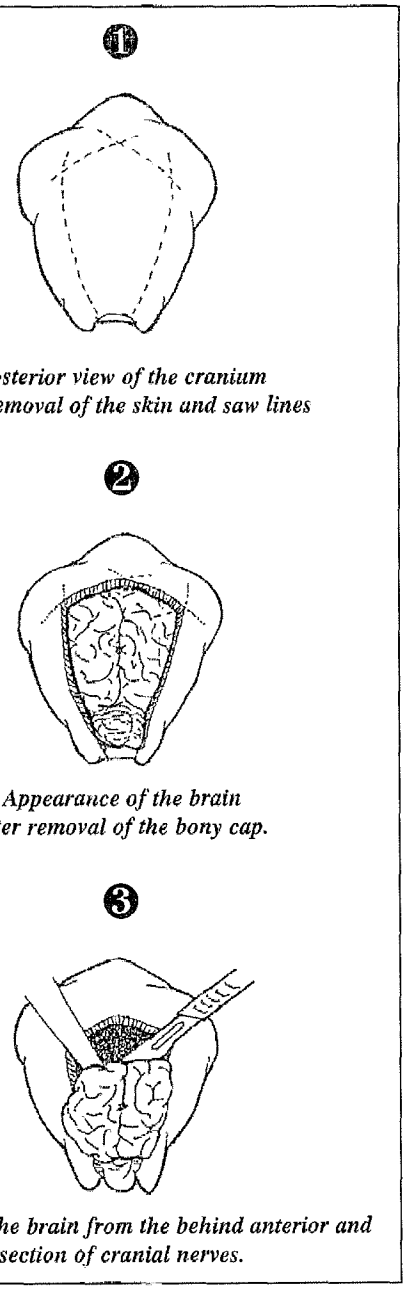

Figure 2.3 : Sample of the brain.

2.4.6 Blood

Samples of blood from ewes that have aborted are the best samples to take at the time of abortion for the isolation of organisms.

Two samples of blood should be taken from each animal in evacuated sterile tubes (e.g. Vacutainer), one tube without anticoagulant for antibody detection, the other with heparin for isolation of organisms.

A second sample of blood should be collected 2 to 3 weeks later in a clean tube without anticoagulant to detect any seroconversion.

When possible, precolostral blood from new born lambs should be sampled at the same time as those taken from the aborted ewes to detect any antibodies to Akabane virus or Border Disease virus.

Blood samples should be collected from about ten females that have aborted. If this number of samples are not collected at the time of intervention, it is possible to make up numbers by collecting samples from females that have not aborted providing that the actual status of the animals is recorded.

If testing is carried out some time after abortions have taken place and if the animals that aborted cannot be identified precisely, samples must be taken from a representative number of adult females or at least twenty of the flock.

The rate of serum antibody production may decrease rapidly depending on the infectious organism involved or the serological technique employed. Blood samples should therefore be collected less than eight weeks after the time of abortion or lambing. If samples are collected beyond this time testing will not always allow a distinction to be made between a latent infection or recent vaccination and the infection responsible for the observed abortions.

REFERENCES

1. Digiacomo R.F., Koepsell T.D. Sampling for detection of infection or disease in animal populations. JAVMA, 1986, 189, 22-23.

2. Eliot M. Méthodologie de dépistage des maladies infectieuses animales. Unité Pédagogique Maladies Contagieuses, Ecoles Nationales Vétérinaires Françaises, Maisons-Alfort. 1992, 1-123.

3. Eliot M., Koutchoukali M.A. Sondage dans une population animale: estimation du taux d'infection des cheptels. Epidemiol Sante Anim, 1984, 6, 65-77.

4. Farver T.B., Thomas C., Edson R.K. An application of sampling theory in animal disease prevalence survey design. Prev Vet Med, 1985, 3, 463-473.

5. Garin-Bastuji B., Mouton F. Méthodologie de l'immunodiagnostic. In: Pastoret P.P., Gouaerts A., Bazin, H. (Eds). Immunologie Animale. Médecine Sciences Flammarion, Paris. 1990, 657-664.

6. Sanaa M., Gerbier G., Eliot M., Toma B. Echantillonnage dans les enquêtes descriptives. Epidémiol Sante Anim, 1994, 25, 45-67.

7. Schwartz D. In: Méthodes Statistiques à l'Usage des Médecins et des Biologistes. Médecine-Sciences Flammarion, Paris. 1983, 1-318.

8. Schwartz D. In: Le Jeu de la Science et du Hasard. Médecine-Sciences Flammarion, Paris. 1994, 1-113.

9. Toma B., Benet J.J., Dufour B., Eliot M., Moutou F., Sanaa M. In: Glossaire d'Epidemiologie Animale. Editions du Point Veterinaire, Maisons-Alfort. 1991, 1-367.