HYBRID STUDY DESIGNS

A **nested case-control study** is similar to a cohort study with the key difference that samples of non-cases are selected for analysis (rather than the entire cohort, as in the case of a cohort study. Nested case-control studies are useful when it is either too costly or not feasible to perform additional analyses on an entire cohort (e.g. if collection of specimens and laboratory analysis of specimens is expensive) Compared with standard case control studies, nested studies:

1) can utilise exposure and confounder data originally collected before the onset of the disease, thus reducing potential recall bias and temporal ambiguity, and

2) include cases and controls drawn from the same cohort, decreasing the likelihood of selection bias.

The nested case-control study is thus considered a strong observational study, comparable to its parent cohort study in the likelihood of an unbiased association between an exposure and an outcome. A concern, usually minor, is that the remaining non diseased persons from whom the controls are selected when it is decided to do the nested study, may not be fully representative of the original cohort due to death or losses to follow-up

A panel study combines the features of cross-sectional and a prospective cohort designs. It can be viewed as a series of cross-sectional studies conducted on the same subjects (the panel) at successive time intervals (sometimes referred to as waves). This design allows investigators to relate changes in one variable to changes in other variables over time.

A repeated survey A repeated survey is a series of cross-sectional studies performed over time on the same study population, but each is sampled independently. Whereas panel studies follow the same individuals from survey to survey, repeated surveys follow the same study population (which may differ in composition from one survey to the next). Repeated surveys are useful for identifying overall trends in health status over time. However, Prospective cohort studies require a long follow-up period. In the case of rare diseases large

groups are necessary. Losses to follow-up can become an important problem. Often it is quite expensive to run.

EXPERIMENTAL STUDIES

The experimental epidemiologist observes and analyses data from groups of animals from which he can select, and in which he can alter, the factors associated with the groups. An important component of the experimental approach is the control of the groups. Experimental studies are also designed to test hypotheses between specific exposures and outcomes. The major difference is that in experimental studies the investigator has direct control over the study conditions.

Randomised clinical trials The randomised clinical trial is the epidemiologic design that most closely resembles a laboratory experiment. The major objective is to test the possible effect of a therapeutic or preventive intervention. The design's key feature is that a formal chance mechanism is used to assign participants to either the treatment or control group. Subjects are then followed over time to measure one or more outcomes, such as the occurrence of disease. All things being equal, results from randomised trials offer a more solid basis for inference of cause and effect than results obtained from any other study design.
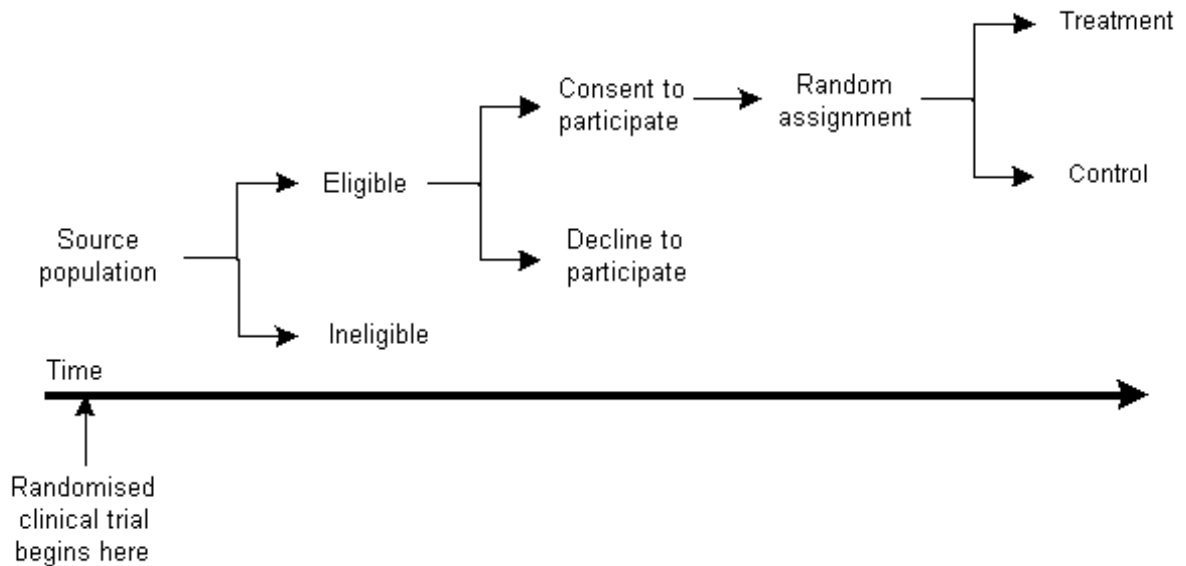
Fig. 3. Schematic diagram of a randomised clinical trial

Advantages: Randomisation generally provides excellent control over confounding, even by factors that may be hard to measure or that may be unknown to the investigator.

Disadvantages: For many exposures it may not be ethical or feasible to conduct a clinical trial (e.g. exposure to pollution). Expensive. Impractical if long periods of follow-up required.

Community trials Instead of randomly assigning individuals to treatment or control groups, community trials assign interventions to entire groups of individuals. In the simplest situation one group (community) receives the treatment and another serves as a control.

QUANTIFICATION OF DISEASE EVENTS IN POPULATIONS

Data used to quantify disease events in populations are often dichotomous in nature i.e. an animal can either be infected with a disease agent or not infected. Such data are frequently presented in the form of an epidemiological *rate.*

In epidemiology, a rate can be defined as the number of individuals having or acquiring a particular characteristic (normally an infection, a disease or a characteristic associated with a disease) during a period of observation, divided by the total number of individuals at risk of

having or acquiring that characteristic during the observation period. The expression is then multiplied by a factor, normally a multiple of 10, to relate it to a specified unit of population.

Rates are commonly expressed as decimals, percentages, or events per standard units of population e.g. per 1000, 10000 animals etc. This produces a standardised measure of disease occurrence and therefore allows comparisons of disease frequencies over time to be made between or within populations. Note that in a rate, the numerator is always included in the denominator, while in a ratio it is not included. In an epidemiological rate, the period of observation should always be defined.

It is difficult to make valid comparisons of disease events between or within populations unless a denominator can be calculated. The use of "dangling numerators" to make comparisons is one of the biggest "crimes" that the epidemiologist can commit, and it should be avoided whenever possible.

For example, suppose we were interested in comparing the numbers of cases of infection with a particular disease agent over a particular time period in two herds of cattle of the same breed but under different management systems. We are told that in herd A the number of animals infected with the disease agent in question in the month of June 1983 was 25, while in herd B the number of animals infected with the same disease agent in the same month was 50. We might therefore conclude, erroneously, that the disease was a greater problem in herd B than in herd A. Note that we did not know the denominator i.e. the population of animals at risk of being infected with the disease agent in each herd. Suppose we investigated further and found that the population at risk in herd A during the month of June was 100 while in herd Bit was 500. Then, calculating a rate for each herd, we find that the rate of infection in herd A was 25/100 or 0.25 or 25% or 250 in 1000, while in herd B it was 50/500 or 0.10 or 10% or 100 in 1000. The true position, therefore, is that the disease was a greater problem in herd A!

The two main types of rates used in Veterinary epidemiology are:

Morbidity rates, which are used to measure the proportion of affected individuals in a population or the risk of an individual in a population of becoming affected.

Mortality rates, which measure the proportion of animals dying in a population.

Morbidity rates

Morbidity rates include incidence, attack, prevalence and proportional morbidity rates.

Incidence rate is the number of new cases of a disease occurring in a specified population during a specified time period, divided by the average number of individuals in that population during the specified time period.

For example, suppose that out of an average population of 4000 cattle in a quarantine camp, 600 animals developed symptoms of rinderpest during the month of June. The incidence of rinderpest in that quarantine camp for the month of June was 600/4000 = 0.15 or 15% or 150 new cases per 1000 animals.

The incidence rate is a way of measuring the risk that a susceptible individual in a population has of contracting a disease during a specified time period. Therefore, if a susceptible animal had been introduced into the quarantine camp on I June, it would have had a 15% chance of contracting rinderpest by the end of the month.

When calculating incidence rates, problems frequently arise in estimating the denominator. Because of births, deaths, sales, movements etc. livestock populations rarely remain stable over periods of time, and such fluctuations in the denominator will obviously affect the calculation of the incidence rate. There are various ways of estimating the denominator in incidence rate calculations. These normally involve measuring the population at various intervals during the study period and averaging the results.

For instance, suppose that in our previous example there were 4000 animals present at the beginning of June but that 100 animals died of the disease by the end of the second week and a further 300 by the end of the month. Assuming that no new animals were introduced or born, the animal population in the quarantine camp at the start of the observation period was

therefore 4000, at the mid-period 3900 and at the end 3600. We might decide to calculate the denominator by taking the populations present at the beginning and end of the observation period and averaging them:

(4000 + 3600)/2 = 3800

The corresponding incidence rate would be 600/3800 = 0.158 or 15.8%.

Alternatively, we might take the populations present at the beginning, middle and end of the observation period and average them -

(4000 + 3900 + 3600)/3 = 3833

- and the incidence rate in this case would be 600/3833 = 0.156 or 15.6%.

Note that the different methods of calculating the denominator have resulted in slightly differing estimates of incidence. Because of this, the method used in calculating the denominator should always be specified when comparisons of incidence are being made, and the same method should be used throughout. Due to difficulties in the calculation of the denominator in incidence rates, another form of morbidity rate, the attack rate, is sometimes used.

The *attack rate* is the total number of cases of a disease occurring in a specified population during a specified time period, divided by the total number of individuals in that population at the start of the specified time period. The denominator, therefore, remains constant throughout the period of observation. Thus, in our previous example, the attack rate would be 600/4000 = 15%.

Strictly speaking, the definition of the attack rate requires that all cases of disease, not just new cases, are included in the numerator. Attack rates are normally used, however, to quantify the progress of a disease during an outbreak. In most instances there would have been no cases of the disease in question prior to the onset of the outbreak, so that all the cases are, in fact, new

cases, and the attack rate becomes a modified form of incidence rate, sometimes referred to as a cumulative incidence rate.

Prevalence rate is the total number of cases of a disease occurring in a specified population at a particular point in time, divided by the total number of individuals in that population present at that point in time.

For example, suppose that in a population of 4000 cattle held at a quarantine camp there were 60 cases of rinderpest when the population was examined on June 18. The prevalence of rinderpest at that camp on 18 June would then be 60/4000 = 0.015 or 1.5% or 15 cases per 1000 animals.

Note that prevalence is a cross-sectional measure referring to the amount of disease present in a population at a particular point in time, hence the term *point prevalence.* However, when dealing with large populations, point prevalence becomes almost impossible to obtain, since it is not possible to examine all the individuals in that population at a particular point in time. In general, therefore, measurements of prevalence have to take place over a period of time, and this is known *as period prevalence.* Provided that the time taken to measure the prevalence remains reasonably short, this parameter retains a fair degree of precision. If, however, the time interval becomes too long, a significant number of new cases of the disease will have occurred since the start of the measurement period. The parameter then becomes a mixture of point prevalence and incidence and, as such, loses precision.

The terms incidence and prevalence are frequently confused and misused. Confusion normally arises due to a failure to define accurately the denominator i.e. the actual population being considered. This can result in the population at risk being either ignored or not considered in its entirety.

Examples of this can be found in reports from veterinary offices laboratories, in which the term "incidence" is often used to express the number of diagnoses or isolations of a particular disease agent as a percentage of the total number of diagnoses or isolations performed. In this

case the denominator is not the population of individuals at risk from the disease, and the rate calculated resembles a form of a proportional morbidity rate.

A proportional morbidity rate is the number of cases of a specific disease in a specified population during a specified time period, divided by the total number of cases of all diseases in that population during that time period.

For example, suppose that an outbreak of contagious bovine pleuropneumonia (CBPP) occurs in a herd of cattle. During a 6-month period there are 45 cases of different diseases, including 18 cases of contagious bovine pleuropneumonia. The proportional morbidity rate for contagious pleuropneumonia in that herd for the 6 months would then be 18/45 = 0.4 or 40% or 400 cases of CBPP in 1000 cases of all diseases.

Mortality rates

The most commonly used mortality rates are crude death rate and cause-specific death rate.

Crude death rate is the total number of deaths occurring in a specified population during a specified time period, divided by the average number of individuals in that population during the specified time period.

The denominator for this rate can be estimated in the same ways as that for an incidence rate. Note, the method of calculating the denominator should always be defined and the same method used throughout to enable meaningful comparisons to be made.

Example: Suppose that in a herd of cattle there were 40 deaths in a year. The number of animals in the herd at the start of the year was 400, at mid-year 420, and at the end of the year 390. The average herd size could therefore be either

(400   +   390)/2   =   395
or
(400 + 420 + 390)/3 = 403

Depending on which method we used to calculate the denominator, the crude death rate would be either 40/395 = 0.101 (10.1%) or 40/403 = 0.099 (9.9%).

Cause-specific death rate is a useful mortality rate and can be defined as the total number of deaths occurring from a specified cause in a specified population during a specified time period, divided by the average number of individuals in that population during that time period. The denominator is calculated in the same way as for an incidence or crude death rate, and the same caveats apply in its calculation.

Example: Suppose that there were 20 deaths from babesiosis in the herd mentioned above, then the death rate due to babesiosis in that herd would be either 20/395- = 0.051 (5.1 %) or 20/403 = 0.050 (5.0 %).

Other useful mortality rates

Proportional mortality rate is the total number of deaths occurring from a specified disease in a specified population during a specified time period, divided by the total number of deaths in that population during that time period.

Example: Suppose that out of 40 deaths in a herd 20 were from babesiosis, then the proportional mortality rate due to that disease would be 20/40 = 0.5 or 50%.

Case fatality rate is the number of deaths from a specified disease in a specified population during a specified period, divided by the number of cases of that disease in that population during that time period.

Example: assuming that there were 50 cases of babesiosis in the herd, then the case fatality rate due to babesiosis would be 20/50 = 0.4 or 40%.

The rates described above are those that are most likely to be used in epidemiological studies in Africa.

The use of specific rates

In epidemiology, we are nearly always involved in studying the effects of determinants on the frequency of occurrence of disease. This often involves the comparison of some of the rates mentioned previously, either in the same population over time - normally before and after a determinant is added or removed - or between populations - either with or without an added

determinant, or with different frequencies of occurrence of the determinant, either at the same point in time or over a period of time.

For such comparisons to be valid, the comparison groups should differ from one another only in the presence, absence, or frequency of occurrence of the particular determinant being studied. Since epidemiology usually involves the study of determinants under uncontrolled field conditions, these criteria are extremely difficult to fulfill. Nevertheless, if rates are expressed in such a form as to ignore the different characteristics which may be present within the disease agents or host populations being compared, there is a danger that such rates may give an oversimplified and even false impression of the actual situation.

Rates can be made more specific, and the comparisons between them more valid, by taking into account various different characteristics. Differences in subspecies and strains of disease agents can be accounted for by clearly defining the subspecies or strain being studied and by making sure that only those individuals affected by that particular subspecies or strain are included in the numerator. Differences in the characteristics of host populations due to age, breed and sex can be expressed by calculating rates **which** take these specific characteristics into consideration.

Thus, for example, one could calculate an *age-specific incidence rate* which is defined as the number of new cases of a disease occurring among individuals of a specified age group in a specified population during a specified time period, divided by the average number of individuals in that specified age group in that population during that time period. Alternatively, one could calculate a *breed-specific incidence rate* which is defined as the total number of new cases of a disease occurring among individuals of a specific breed in a specified population during a specified time period, divided by the average number of individuals of that breed in that population during that time period. One could go even further and calculate an *age-breed specific incidence rate* which is defined as the total number of new cases of a disease occurring among individuals in a specified age group of a specified breed in a specified population,

divided by the average number of individuals of that specific age and breed in that population during that time period.

The same procedures can be applied to other morbidity and mortality rates. A large variety of specific rates can thus be calculated by using appropriate definitions of the numerator and the denominator. As a general principle, rates should be made as specific as the data allow, but not so specific as to make the numbers involved too small for statistical analysis. For analytical purposes there is little or no advantage in calculating and comparing age- or breed-specific rates if an age-breed specific rate can be calculated.

The following is an example illustrating the advantages of using specific rates in making comparisons. Suppose we wished to assess the efficiency of a tick control programme in two East Coast fever (ECF) endemic areas, where the level of disease challenge, the environmental conditions and the systems of management were approximately the same. In area A there was an average population of 10 000 head of cattle present during a 1-month study period, and 500 animals from that population developed symptoms of ECF during that period. In area B there was an average population of 15 000 head of which 1500 developed symptoms of the disease during the study period. The crude incidence rate of the disease in area A was 500/10 000 = 5 % and in area B 1500/15 000 = 10%. We might conclude, therefore, that the tick control programme in area A was more efficient than in area B.