11

CONTROLLED STUDIES

OBJECTIVES

After reading this chapter, you should be able to:

1. Design a controlled trial to produce a valid evaluation of an intervention, paying special attention to:
    a. the statement of objectives of the trial
    b. the definition of the study subjects
    c. the allocation of subjects to the interventions
    d. the identification and definition of appropriate outcome variables
    e. ethical considerations in the design and implementation of the trial.
2. Conduct a controlled trial efficiently, while paying special attention to:
    a. masking as a procedure to reduce bias
    b. following all intervention groups adequately and equally
    c. developing and using appropriate data-collection methods and instruments
    d. proper assessment of the outcomes being measured
    e. correct analysis and interpretation of the results
    f. clear reporting of methods and results.
3. Design and conduct a valid controlled trial of a vaccine, or prophylactic, against an infectious agent.

11.1 INTRODUCTION

A controlled trial is a planned experiment carried out on subjects in their usual environment. Particular care must be taken in the design and execution of these studies because they often involve client-owned animals, and their size and scope make it very difficult to replicate them for the purpose of validating the findings.

Controlled trials are especially useful for the evaluation of interventions that can be easily manipulated such as therapeutic or prophylactic products, diagnostic procedures and animal-health programmes. Most trials are conducted to assess one specific intervention and, indeed, this is their forte. The outcome might include a specific health parameter (*eg* clinical disease), or a measure of productivity, performance or longevity. The study groups are formed by random assignment of the intervention(s) being evaluated and can be composed of individual animals, herds or other groups. Lavori and Kelsey (2002) edited a comprehensive review of clinical trials; this provides an excellent overview of trial design, analysis and interpretation. A special issue of Statistics in Medicine (Vol 21, Issue 19, 2002) was devoted to a discussion of long-term clinical trials.

The term **clinical trial** is often used synonymously for controlled trial. However, some authors restrict its use to trials of therapeutic products and/or trials carried out in a clinical setting. We will use the term 'controlled trial' to refer to planned experiments designed to evaluate products or procedures in subjects outside the laboratory. Because controlled trials can be used to investigate a wide range of products/programmes, we will refer to the factor being investigated (*eg* treatment) as the **intervention**, and to the effect of interest as the **outcome**. Animals, or groups of animals participating in the trial will be referred to as **subjects** (regardless of whether they are individual animals, herds or other populations of animals). Animal owners will be referred to as **participants**.

Controlled trials are, by far, the best way for evaluating animal-health interventions because they allow much better control of potential confounders than observational studies, as well as reducing bias due to selection and misinformation (..."the randomised controlled trial is at present the unchallenged source of the highest standard of evidence used to guide clinical decision-making" Lavori and Kelsey, 2002). In the absence of evidence as to the efficacy and safety of animal-health products and procedures derived from controlled trials, practitioners would be left in the unenviable position of making decisions about their use based on extrapolation of data from studies carried out under artificial (laboratory) conditions or based on their own limited and uncontrolled experience. Having said this, the results of many trials have been criticised for being "of limited relevance to answering questions about whether an intervention does work under usual circumstances" (Zwarenstein *et al*, 2006). This issue has lead others to describe how to design trials to investigate practical problems (Treweek *et al*, 2006) and to develop a specific 'tool' to help researchers prepare high-quality research proposals for clinical trials.

11.1.1 Phases of clinical research

While controlled trials are valuable for assessing a wide range of factors affecting animal health and productivity, one of their most common uses is to evaluate pharmacological products (therapeutic and preventive). Consequently, a brief review of the phases of research used in the development and evaluation of these products is warranted.

Clinical pharmaceutical research can be divided into 4 phases.

- **Phase I** trials (sometimes referred to as formulation trials) are studies carried out in healthy animals primarily to evaluate safety of the drug (*eg* to determine safe dosage ranges, or identify adverse reactions *etc*).
- **Phase II** trials are the first evaluation of the drug in a small number of animals from the target population (*eg* sick animals). They are used to document the activity of the drug. These studies might involve before/after comparisons and often there is no specific control group.
- **Phase III** trials are large-scale experimental studies to determine the efficacy of a drug in a typical clinical population, to monitor side effects and to compare the drug with other available treatments. These studies should be based on randomised controlled trials. While generally required to be carried out before the registration of products for human use, they are not necessarily required for registration of animal-health products in all countries. These studies need to be carried out according to good clinical practice (GCP) standards. GCP is a standard for designing, conducting, monitoring, recording, auditing, analysing and reporting clinical studies. A set of standards, developed under the International Cooperation on Harmonization of Technical Requirements for Registration of Veterinary Medicinal Products, are available (http://www.vichsec.org/).
- **Phase IV** trials are post-registration trials designed to evaluate the most effective way of using a product. They also should be carried out as randomised controlled trials, although they require less documentation than studies used in the product registration process. In the absence of randomised controlled trials carried out prior to registration, they provide the most reliable information about the efficacy of a product in the context of everyday real-world activities.

11.1.2 Key design elements

An important feature in the design of a controlled trial is the development of a detailed study protocol which covers all elements of the study design and execution. This 'road map' includes: stating the objectives, defining the source population in which the study will be conducted, allocation of subjects, specifying the intervention, masking (blinding), follow-up and compliance, specifying and measuring the outcome, analysis of trial results, and ethical considerations. These aspects of the trial design are related to the features that should be reported when describing the results of a trial (see Section 11.12). Each of these will be considered in this chapter. A number of specific articles are referenced, as examples, throughout this chapter.

11.2 STATING THE OBJECTIVES

The objective(s) of the trial must be stated clearly and succinctly. This explicit statement should describe the intervention being investigated and the primary outcome(s) to be measured; the objective should also allude to the unit of concern. As a general rule, each trial should have a limited number of objectives (see Example 11.1 for a straightforward trial) with 1 or 2 primary outcomes; some trials might also include a small number of secondary outcomes (see Example 11.9). Increasing the number of objectives may unnecessarily complicate the protocol and might jeopardise compliance and other aspects of the trial. A trial with a very simple design might be able to include a much larger sample size within a given budget, thus enhancing the

Example 11.1 Eprinomectin treatment of psoroptic mange in hunter/jumper and dressage horses: a prospective, randomised, double-blind, placebo-controlled clinical trial

The objective of this clinical trial was to investigate the efficacy of topical eprinomectin for the treatment of psoroptic mange infestation in horses (Ural *et al*, 2008). Twenty four privately owned hunter/jumper and dressage horses were diagnosed with psoroptic mange infestation based on clinical findings and parasitological skin scraping results. Each horse was randomly assigned to either the topical eprinomectin pour-on solution (at a dose of 500 μg/kg body weight, once weekly for 4 applications) treatment group or to a placebo group (distilled water; we do have a concern that both the owner and the veterinarian could differentiate between water and the eprinomectin solution). Clinical evaluations and skin scrapings were done by the same veterinary investigator at the beginning, during and at the end of the treatment. Both owners and the veterinary investigator were blinded to the allocation to the groups. Horses were examined for psoroptic mites (recorded as present or absent) on days 7, 14, 21, 28 and 40 for follow-up. Fisher's exact test was used to assess differences in the number of horses without mites on each assessment date between the eprinomectin treatment and placebo groups.

power of the study.

This chapter will focus on controlled trials that contrast 2 groups (sometimes referred to as 2-arm studies) (*eg* Example 11.1), although the principles also apply to studies with more than 2 'arms'. The latter may require a more complex design, and a larger sample size, although some efficiency in this area can be obtained through the use of factorial designs (see Section 11.4.2). The 2 groups might be a comparison of an intervention with a placebo, no treatment, the usual treatment, or a different dose of the same product. The trial can be active (concomitant groups) or historical for the control group (D'Agostino *et al*, 2006). Placebos are ideal when there is no established alternative intervention, and where possible, a placebo should be used in preference to 'no treatment'. However, it is unethical to include a placebo or no-treatment group if it will result in undue suffering of animals that they wouldn't experience under appropriate management (*ie* an existing product or procedure exists to reduce or prevent that suffering). Furthermore, the decision as to whether to use a positive control (existing therapy) or negative control (placebo) might have profound effects on the subjects available for inclusion in a trial, as well as the results. For example, farmers, knowing that there was a 50% chance that any animals they enrol in the trial would receive no useful treatment, might only be willing to submit animals which they were already planning on culling. Since the 'control' treatment level is often the current standard treatment, a frequently used trial design is the non-inferiority trial. This is an active controlled trial to investigate whether a new intervention is at least not inferior to the existing best intervention. An introduction to the design of these trials has been published recently (D'Agostino *et al*, 2006).

11.3 THE STUDY GROUP

When designing a trial, you should be able to specify the target and source populations. The target population is the population to which you want the results of the trial to apply (see Chapter 2), and is always an important feature, especially in phase III clinical trials in which the geographic location could play a role in the acceptability of the trial for the registration process. The source population should be representative of the target population and represents the subjects who are eligible for the trial. Finally, the study group is the collection of subjects in

which the trial will be carried out. If the study group is not randomly obtained from the source population, it should be representative of it. Usually, the study group is obtained by seeking volunteer participants either by contacting them directly (*eg* personally (see Example 11.1), via letter, or the media) or by asking veterinarians to nominate some of their clients whose subjects meet the eligibility criteria. While the use of volunteer participants is unavoidable, how well the study group (participants or study subjects) represent the source and target populations (see Section 2.1.3) must be taken into consideration when extrapolating the study results.

11.3.1 Unit of concern

When defining the source population, the first issue is to specify the level of organisation at which the intervention will be applied (*eg* individuals as in Example 11.1, or cages of fish as in Example 11.2). If an intervention can only be applied at a group level (*eg* to a litter, pen or herd) then the unit of concern consists of groups (*eg* all enrolled sea cages on the study farms in the teflubenzuron trial; Examples 11.2 and 11.3). The outcome in such a study might be measured at the group level (a group-level study) or at the individual level (a cluster randomised study—discussed in Section 11.4.2). The same applies if the intervention is applied at the individual level but the study is designed so that all individuals within a group receive the same intervention.

11.3.2 Eligibility criteria

Once it has been determined whether individuals or groups will be recruited for the study, eligibility criteria need to be considered using some, or all, of the following factors:

- Animal-handling facilities and personnel must be in place to allow for the necessary sampling during the trial.
- Adequate records must be available to document the subject's past history and to provide outcome measures (if relevant).
- For trials of therapeutic agents, clear case definitions for the disease being treated must be developed to determine which cases are eligible for inclusion.
- For trials of prophylactic agents, healthy subjects are required and procedures for documenting their health status at the start of the trial might be required.
- Subjects in a trial need to be capable of benefiting from the intervention. As much as possible, avoid the 'ceiling effect' (the maximum possible improvement). For example, the start of the teflubenzuron trial (Example 11.2) was delayed by a month because the

Example 11.2 Randomised clinical trial to investigate the effectiveness of teflubenzuron for treating sea lice on Atlantic salmon

A double-blind, randomised controlled trial was performed to investigate the effectiveness of teflubenzuron in controlling sea lice *Lepeophtheirus salmonis* on farmed Atlantic salmon *Salmo salar* (Campbell *et al*, 2006b). Cages were the unit of concern, and 40 sea cages from 3 commercial cage sites in Atlantic Canada were used in this trial. Pairs of cages were matched by site, cage size, and pre-treatment mean lice counts and then randomly assigned to receive medicated feed or not. The teflubenzuron was administered in the feed at a dosage of 10 mg/kg biomass per day for 7 days. Post-treatment lice counts and staging of developmental stages were performed at 1 and 2 weeks after the end of treatment. Linear regression, using log transformed counts of sea-lice per fish with cage as a random effect was used for analysis.

Example 11.3 Historical control clinical trial to assess the effectiveness of teflubenzuron for treating sea lice on Atlantic salmon

A historical controlled trial was performed to assess the effectiveness of teflubenzuron in controlling sea lice *Lepeoptheirus salmonis* burdens on farmed Atlantic salmon *Salmo salar* over time (Campbell *et al*, 2006a). The study site comprised 9 sea cages, all of which were treated. The teflubenzuron was administered in the feed, at a dosage of 10 mg/kg biomass per day, over a treatment period of seven days. The effectiveness of the intervention was assessed at weekly intervals by comparing pre-and post-treatment lice counts, linear regression, using log transformed counts of the differences in sea lice per fish between the initial and subsequent samplings, with cage as a random effect used for analysis.

general level of sea lice in the Bay of Fundy was slow to build during the summer the trial was carried out. There was no point evaluating the intervention when there were too few lice to work on. Restriction of a trial to subjects that are most likely to benefit from the intervention will increase the power of the trial but might limit the generalisability of the results.

- Avoid subjects with high risks for adverse effects.

Eligibility criteria, must be stated clearly, and applied to the study subjects; these can range from individual privately owned animals (Example 11.4), animals within a herd (Example 11.5), to groups of animals (*eg* cages within a fish farm; Example 11.2). In some instances, if the participants do not meet the selection criteria at the time of recruitment (*eg* not having adequate records), it might be acceptable to have them agree to meet the standards during the period of the trial. A narrow set of eligibility criteria will result in a more homogenous response to the intervention and this might increase the statistical power of the study, but reduce the generalisability of the results. A broad set of eligibility criteria will result in a much larger pool of potential applicants, but there could be a large background variation in study subjects (this can be advantageous for detecting variation in response to the intervention in subgroups of subjects), but can have negative effects on the power of the trial. Balancing these 2 considerations must be done on a case-by-case basis, while adhering to the objectives of the study. In general we suggest using eligibility criteria that reflect the breadth of subjects who

Example 11.4 Stratification, blinding and placebo effect in a randomised, double blind placebo-controlled clinical trial of gold bead implantation in dogs with hip dysplasia

Eighty dogs with canine hip dysplasia (CHD) were included in a randomised, placebo-controlled and double-blind clinical trial with stratified parallel group design (Jaeger *et al*, 2005). Body weight (3 groups) and degree of CHD (2 groups) were used as stratification factors. Dogs with other diseases related to the nervous, muscular or skeletal systems were excluded. Thirty-eight dogs were allocated to gold bead implantation and 42 to placebo. After 6 months, 33 of the 42 placebo-treated dogs received gold bead implantation in an open study lasting a further 18 months. The main outcome variable in the study was change in pain from CHD as assessed by the owner. Losses from the study were carefully documented. No significant difference in the main outcome variable, regardless of the treatment given, could be detected.

Owners were asked to guess the intervention their dog received and 60% of the owners correctly guessed the intervention; this was not statistically significantly different from the expected 50%. However, among those that guessed that their dogs received the gold beads, the improvement was judged to be higher than in the group who thought (or were uncertain) their dog received the placebo The authors concluded that a significant placebo effect on the evaluation of the treatment was likely.

Example 11.5 A clinical trial evaluating prophylactic and therapeutic antibiotic use on health and performance of preweaned calves

The objective of this clinical trial was to evaluate the influence of prophylactic and therapeutic antibiotics on health and performance in pre-weaned multisourced dairy calves on a calf ranch (Berge *et al*, 2005). 120, one-day-old calves were enrolled, over a 2-day period and allocated into 1 of 3 management systems for antibiotic use. The outcome was morbidity over a 4 week period and was recorded as the number of sick days. A morbidity reduction of 1.5 days (SD = 2 days) was used to determine the sample size; an additional 15% was added for potential losses. Calves were allocated to groups in a systematic manner as they were unloaded from the truck. Sixty calves received no antibiotics; 30 calves were eligible to receive individual antibiotic treatment for disease, but no prophylactic antibiotics in milk replacer; and the remaining 30 calves received milk replacer medicated with neomycin and tetracycline HCl, and could also be treated therapeutically with antibiotics. Health status and treatments were monitored and recorded daily. Managers were blinded as to study group. Three primary study outcomes were weight gain, morbidity, and mortality.

might receive the intervention in the future if it is shown to be effective (Zwarenstein *et al*, 2006).

11.3.3 Sample size

We begin this discussion assuming that we are designing a trial with a fixed sample size, the most common approach used for clinical trials. The size of the study needs to be determined through appropriate sample size calculations (see Chapter 2), with attention paid to the estimated effect of the intervention and both Type I and Type II errors. The magnitude of the effect to be detected (or estimated) should be clinically (and in some cases, economically) meaningful. When computing the power of the study (1-Type II error), it is common to set the power to 90%. **Note** The sample size needs to be increased if you want to carry out meaningful investigations of the effect of the intervention in subgroups of the study population.

As has been noted, the sample size required for qualitative (*eg* dichotomous) outcomes is often much larger than that required for outcomes measured on a continuous scale. Obviously, the choice of outcome(s) and its measurement should reflect the study objectives. The basic formulae for sample-size calculation, where the individual subject is the unit of randomisation, and the outcome is either binary or continuous, are presented in Chapter 2. Here we mention a few important issues that impact on sample size. Auleley *et al* (2004) discuss planning the sample size and how it is impacted by choice of the outcome measure(s), the scale of the outcome (*ie* continuous, binary, or time to event), and the occurrence of missing values. Barthel *et al* (2006) also discuss sample size issues when the outcome is time to event (*ie* survival), and provide a very flexible program in Stata known as ART (Analysis of Resources for Trials) for planning sample size in complex designs that allows for adjustments for missing data, non-proportional hazards and censoring (see Chapter 19). If there are multiple outcomes of approximately equal merit, using a generalised approach to power based on the probability that important changes in all outcomes will be observed has been described by Borm *et al* (2007a). Korn and Freidlin (2006) update the correct approach to determining sample size if historical controls will be used.

Sample size for the allocation of clusters of subjects

Cluster randomised trials are those in which all subjects within a group (*eg* cows in a herd) are

allocated to the same intervention (see Section 11.4.2). In planning such trials, we need to account for the intra-cluster correlation (ρ) and the cluster size (m). As noted in Section 2.11.6, the sample size for a study needs to be increased by a factor of $(1 + \rho(m\text{-}1))$, so even if ρ is small, the overall sample size can become very large if the cluster size is large. With respect to cluster size, it has been shown that the power of a study does not increase appreciably once cluster size exceeds $1/\rho$ (Campbell *et al*, 2006b).

When trying to decide on the unit (*ie* individual or group and size of group) to randomise, as the intra-cluster correlation coefficient increases, random allocation of individuals becomes much more efficient statistically, especially if the cluster size is large. However, if the intervention is allocated to clusters, and the number of clusters available is small, a matched design (*eg* matching on strong cluster-level confounders) may be used. When feasible, employing a cluster cross-over design can add to the efficiency particularly when the number of clusters is small (Turner *et al*, 2007).

Sample size for sequential and adaptive designs

A **sequential design** trial incorporates 'a method allowing hypothesis tests to be conducted on a number of occasions as data accumulate through the course of a trial' (Todd, 2007). Thus, the sample size is not fixed in advance of the trial, rather sequential designs have specified stopping rules. Typically the planning of these studies is more complex than the fixed trial design and there is the potential for bias in that the researchers might alter the implementation of the trial after learning the results of the interim analyses. Zou *et al* (2005) describe sequential methods for cluster randomisation.

Adaptive design studies are ones in which the design may change as the study progresses. Consequently, they are more flexible than sequential designs (Golub, 2006). The most common 'adaptation' is modification of the sample size of the second stage based on the predicted power of the trial at the end of the first stage. However, adaptive designs also include dropping or adding treatment arms, changing the primary endpoint, and even changing objectives (for example, switching from non-inferiority to superiority) (Todd, 2007). Outcome adaptive designs strive to ensure that the majority of subjects get the benefit of the best therapy available. The allocation of subjects is influenced by the experience of previous subjects in the trial. One example is 'play the winner' allocation in which subjects continue to be allocated to an intervention level as long as that treatment is producing beneficial results. As soon as it fails, the allocation switches to the other treatment. These procedures are only suitable if the result of the intervention is clearly identifiable in a very short period after treatment. These designs have not been used commonly in animal-health studies.

Other sample size issues

Another issue to consider when planning the size of the trial is the time to recruit study subjects (2 days in Example 11.5; 5 months in Example 11.6). The length of time it will take to recruit the required number of study subjects can be a serious problem for studies on therapies for relatively rare conditions. Two other specific issues related to time for recruitment deserve consideration. First, if recruitment on a study farm lasts longer than one production cycle (*eg* intercalving interval in dairy herds), then an intervention that is related to the production cycle might be reapplied to cows that have already been treated. This might or might not be acceptable (depending on the nature of the intervention), but at the very least will require special consideration in the analyses. Second, if season of treatment is likely to influence the results, then the recruitment period should span at least one full calendar year.

Example 11.6 A two-dose regimen of a vaccine against *Escherichia coli* O157:H7 type III secreted proteins reduced environmental transmission of the agent in a large-scale commercial beef feedlot clinical trial

A clinical vaccine trial of commercially fed cattle tested the effect of a 2-dose regimen of a vaccine targeted against Type III secreted proteins of enterohemorrhagic *Escherichia coli* O157:H7 on the probability of detecting the organism on environmental sampling devices (Smith *et al*, 2008). Nineteen commercial feedlots, provided a total of 70 pairs of pens (vaccinated and non-vaccinated 'usually' (we would recommend 'always') allocated randomly) matched by reprocessing schedule and time of sampling. Recruitment took about 5 months. Vaccine was administered to all cattle within treated pens at arrival processing and again at re-implant processing. Pens of cattle were sampled one week after the second dose of vaccine and every 3 weeks thereafter for 4 test periods. Pair-matched pens of cattle were sampled concurrently. Test samples were 7 ropes per pen hung overnight from the feed-bunk neck-rail. Recovery of *E. coli* O157:H7 from at least one rope classified a pen as positive. The probability of a pen to test positive was modelled using multilevel logistic regression with a random effect for pen.

Loss of subjects from the study might happen for a variety of reasons and can lead to bias. Some subjects might be lost to follow up (*eg* owner moved away or identification tag lost) while others might be non-compliers (*eg* participants who do not comply with the protocol). Finally, some subjects might be lost due to competing risks (*eg* die from other diseases while still on the trial). Once a sample size has been estimated, it is wise to compute the expected power of the study based on different estimates of the potential losses to the study, and adjust the sample size accordingly.

11.4 ALLOCATION OF STUDY SUBJECTS

It is important to remember that controlled trials are based on volunteer subjects (or at least a participant has volunteered them), and participants must agree to have their subjects receive either of the interventions as determined by the allocation process. Once a subject has been enrolled, the allocation should be carried out close to the time at which their participation in the study is scheduled to start.

It is clear that a formal randomisation process is the best method for allocating subjects to study groups; indeed, without this formal allocation procedure, bias is very likely to distort the findings (Gluud, 2006). Using clinical judgement in the selection of interventions can build bias into a non-randomised trial (clinical judgement applied in selection of therapies will likely mean that confounders are unevenly distributed across study groups and hence, bias the trial). It is very difficult to control this bias analytically. The use of propensity scores (Section 13.5) is one potential approach to dealing with this problem. However, propensity scores are generally used in observational studies of interventions in situations in which a randomised control trial (RCT) is not feasible. However, before discussing formal randomisation procedures we will discuss some alternatives.

11.4.1 Alternatives to randomisation

Historical control trials are ones in which the outcome after an intervention is compared with the level of the outcome before the intervention (before/after comparison). For example, a vaccine for neonatal diarrhea might be introduced into a dairy herd and the incidence of

diarrhea in the year after vaccination compared with the incidence in the year before. However, many factors could bias the results in this design and hence, historical control trials are generally unacceptable. For a historical control trial to have any validity, 4 criteria must be met:

1. the outcome being measured must be predictable,
2. there must be complete and accurate databases on the disease of interest,
3. there must be constant and specific diagnostic criteria for the outcome, and
4. there must be no changes in the environment or management of the subjects in the study.

Rarely are many, let alone all, of these criteria met for animal-health problems, although a study of interventions to change the progression of vaccine-induced sarcomas is perhaps one such example (Rassnick *et al*, 2006). An additional limitation of historical control trials is that it is impossible to use blinding techniques. In Example 11.3, given the very short duration of the trial and the investigator's control over all pre- and post-treatment data, a historical control design was considered to be acceptable.

Systematic assignment of individuals to treatment groups (*eg* alternating assignment) can be a reasonable alternative to formal randomisation under field conditions. Systematic assignment might be based on the use of pre-existing animal identification numbers with odd and even numbers forming the basis of the group assignment, or it could be based on the order of study subjects (*eg* animals passing through a chute—see Example 11.5). Systematic assignment might make it harder to keep participants and study personnel blind as to the intervention identity, but aside from this, it is often just as effective as random allocation (provided outcome assessment is done blindly). If half the subjects are to be allocated to receive the treatment, the initial subject allocation should be random and thereafter, every second subject would receive the allocated intervention. Do not apply the intervention to the first (or last) half of the subjects and the comparison treatment to the remainder.

11.4.2 Random allocation

As indicated, formal randomisation is the preferred method of allocation. It must be noted that random allocation does not mean 'haphazard' allocation and a formal process for generating random intervention assignments (*eg* computer-based random-number generator, or even a coin toss—see Example 11.8) must be employed. Random allocation should be carried out as close as possible to the start of the study to reduce the possibility of withdrawals after allocation.

Simple randomisation involves each subject being assigned to an intervention level (*eg* vaccine or not, treated or not) through a simple random process without any further considerations. Stratified randomisation (*eg* randomisation within age categories) helps ensure that a potential confounder (age) is equally distributed across study groups. One specific form of stratified randomisation is random allocation of animals within blocks (*eg* every X animals) to ensure temporal balancing of intervention allocation. (In Example 11.4, dogs were allocated within blocks of 4). This ensures that all temporal or block-level factors that might influence the outcome are balanced across study groups. Hofmeijer *et al* (2008) propose a method of block allocation to gain efficiency in small trials. It adjusts the assignment of the next subject depending on the imbalance in treatment allocation that exists at the time.

Cross-over studies

In a cross-over study, each subject gets both of the interventions (in sequence). However, the first intervention administered is still assigned randomly. This process is only suitable for the

evaluation of therapies where the condition of the subject is stable and the duration of the intervention effect is relatively short-lived. A 'wash-out' period might be required between interventions. It has the advantage that it increases the power of the study since the same subject receives both levels of the intervention. A cross-over trial to study the effect of ionophore treatment (evaluating the effect of monensin) on fecal shedding of Mycobacterium avium subsp. paratuberculosis (Map) has been reported (Hendrick *et al*, 2006). Chronically infected cows were treated with either ionophore or a placebo for 3 months followed by a wash-out period (one month), after which the cows were switched to the other treatment.

Factorial designs

This design is particularly well-suited to trials investigating 2 or more interventions, especially if the interventions might produce synergism or antagonism. Here all possible combinations of the treatments (*eg* neither treatment, treatment 1 only, treatment 2 only, both treatments) are assigned to the study subjects. Because the design is usually balanced, the treatment effects are not confounded (*ie* they are unrelated, or orthogonal, to the intervention) and the analyses are straightforward. Normally, one should not attempt to assess more than 2-3 interventions as the possible interactions become difficult to interpret. Silva *et al*, 2005 report on a factorial design in which the topical action of sodium hypochlorite with and without the systemic use of oxytetracycline for the treatment of clinically diagnosed bovine digital dermatitis was investigated.

Cluster randomisation

There are a number of reasons why a cluster of animals (*eg* a herd) should be allocated to an intervention group rather than individual animals. In some instances, it might be the only feasible method. For example, if the intervention is one which is always given at the group level (*eg* medication in the drinking water), then there is no choice. Even if the intervention could be administered at the individual level, it might be impossible to keep track of individuals within the group, or the intervention in some subjects could influence events (*eg* through spread of a live vaccine) in non-intervention subjects housed with them (see Section 11.10) so assignment of the whole group to one intervention would be appropriate. Cluster randomisation is also appropriate if there is potential for physical spread of a treatment to the control group (*eg* pour-on endectocides when applied to half the cattle in a herd), or the potential for the effects of the intervention to impact the non-intervention groups as in herd immunity (see Example 11.6). One of the largest cluster-randomised trials in veterinary medicine is described in Example 11.7. Recent developments in the design and analysis of cluster randomised trials has been reviewed (Campbell *et al*, 2007).

Cluster randomised trials are much less statistically efficient than trials with random allocation of individuals and the clustering of individual subjects within the groups needs to be taken into account in analysis (see Chapters 20-23). In a cluster randomised trial, the best scenario for follow-up is if all individuals can be monitored for the duration of the study. If this is not possible, following a randomly selected cohort would be the most statistically powerful approach. If it is not possible to follow individuals, the investigator will have to carry out repeated cross-sectional samplings throughout the follow-up period (Campbell *et al*, 2007). Donner and Klar (2004) review the advantages and pitfalls of using cluster randomisation; Donner *et al* (2007) also comment on 'breaking the matches' to gain some statistical efficiency in the analysis of matched-cluster randomised trials.

Example 11.7 A large cluster randomised trial of cattle tuberculosis control based on culling of badgers in the United Kingdom

The randomised badger control trial (RBCT) was launched in 1998 to evaluate the effectiveness of badger culling as a control strategy for cattle tuberculosis (TB) in Britain. The RBCT involved comparing the incidence of cattle TB under 3 experimental treatments; repeated widespread ('proactive') culling of badgers, localised ('reactive') badger culling, and no culling ('survey only'), each replicated ten times in large (100 km^2) trial areas recruited as matched sets of 3, known as 'triplets' (Donnelly *et al*, 2003). The triplets were located in areas of high cattle TB incidence. All trial areas were randomly allocated to treatments (except in one triplet where security concerns directed a specific allocation). Cattle TB was detected by routine testing under the national control scheme.

Split-plot designs

A final elaboration of allocation discussed here is a split-plot. This design is used if there are 2 or more interventions, one of which needs to be applied at the group level and the other(s) can be assigned to individuals. The analysis must take account of the different degrees of freedom to assess intervention effects at the different levels (*ie* group versus individual). (Ritter *et al*, 2009), combined a spilt-plot and factorial design when investigating the effect of stressors on market weight pigs. Density of animals during transport was the whole-plot factor with movement distance and handling intensity being individual pig-level (*ie* split-plot) factors.

Multicentre trials

If an adequate number of subjects is not available at a single site, a multicentre trial might have to be planned (Fedorov & Jones, 2005). A key feature is that the within-centre and between-centre variances need to be accounted for. Although a multicentre trial complicates the protocol and the implementation of the trial, it can enhance the generalisability of the results (because of the usually larger geographic area covered by the trial) and also increases the opportunity to identify interaction effects (*eg* different responses by centre). One key for statistical efficiency in multicentre trials is to try and maintain approximately the same number of subjects per centre (Dragalin & Fedorov, 2006). Example 11.8 describes a multicentre trial of an anthelmintic in dairy cows.

Example 11.8 Effect of eprinomectin pour-on treatment around calving on reproduction parameters in adult dairy cows

The objective of this study was to investigate if treatment of cows with eprinomectin around calving had any beneficial effects on the calving to first artificial insemination interval, calving to conception interval, and number of services per conception in totally and semi-confined dairy herds (Sithole *et al*, 2006).The study was carried out between February 2002 and February 2003 in 35 herds (2,381 cows) participating in a larger clinical trial and located in Quebec, Ontario and Minnesota (USA). The herds kept electronic reproduction records. Cows were randomly allocated to receive eprinomectin or a placebo (mineral oil in bottles matching the eprinomectin), with treatment being administered on, or close to, the day of calving. Monthly bulk tank milk samples from each farm were tested with an indirect ELISA using a crude *Ostertagia ostertagi* antigen and these data were averaged over the study year. The optical density ratio (ODR) values were then dichotomised into high and low using a cutpoint of 0.50. Treatment effects on calving to conception and calving to first service intervals. were analysed using Cox proportional hazards survival models. A random effects Poisson regression model was used to model the number of services per conception.

11.5 SPECIFYING THE INTERVENTION

The nature of the intervention must be clearly defined. A fixed intervention (one with no flexibility) is appropriate for assessing new products (particularly in phase III trials). A more flexible protocol might be appropriate for products that have been in use for some time and for which a body of clinically applied information exists. For example, feedlot cattle might be assigned to 1 of 2 antibiotics for the treatment of respiratory disease but the timing of a decision to change antibiotics (*ie* a treatment failure) or stop treatment (*ie* a treatment success) might be left up to the person responsible for the animals provided it fell within a range defined in the protocol (*eg* between 3 and 5 days). When possible, the initial treatment assignment should remain masked so that clinical decisions are not influenced by knowledge of group allocation. Clear instructions about how the intervention needs to be administered, or implemented, are essential, particularly if participants are going to be responsible for some or all of the interventions. In addition, the system of ensuring that the correct treatment goes to the right animal must be kept as simple as possible and a method of monitoring the intervention administration process should be put in place.

11.6 MASKING (BLINDING)

A key component in the effort to prevent bias in controlled trials is the use of masking (or blinding). Unfortunately, the usage of the terms single, double and triple blinding is not consistent. For our purposes, a single-blind study means that the participant is unaware of the identity of the intervention applied to individual study subjects. This feature should help ensure equal follow-up and management of subjects in the various intervention levels. A double-blind study means that both the participant and selected members of the study team (*ie* people administering the interventions and those assessing the outcomes) are unaware of intervention assignment. This feature helps ensure equal assessment of the subjects in different intervention levels. In a triple-blind study, those who are analysing the data also are unaware as to which group received which treatment. This feature is designed to ensure that the analysis is conducted in an unbiased manner. It is recommended that the success of blinding be evaluated and not taken for granted (Boutron *et al*, 2005). Hrobjartsson *et al* (2007) discuss some approaches to this.

In many cases it is necessary to use a placebo to ensure that the relevant individuals remain blind. A placebo is a product that is indistinguishable from the product being evaluated and which is administered to animals in the groups designated to receive the comparison treatment. In many drug trials, the placebo is simply the vehicle used for the drug, but without any active ingredient (see Example 11.9). One concern with the use of a placebo is that, even though it might not contain the active ingredient being investigated, it could still have either a positive or negative effect on the study subjects. For example, a placebo vaccine that does not contain the antigen of interest might still induce some immunity as a result of adjuvant in the placebo. These issues should be discussed and settled prior to conducting the trial.

In some cases, using a placebo might not be adequate to ensure blinding. For example, in trials of recombinant bovine somatotropin (rBST) in dairy cattle, it has been argued that a placebo is irrelevant because the drug produces such a noticeable change in milk production, anyone working with the cows on a regular basis would know which cows received the treatment. Nonetheless, masking the intervention should be used whenever possible.

Example 11.9 Effects of a commercially available vaccine against *Salmonella enterica* serotype Newport

180 convenience sampled non-pregnant cows were selected on a 1,200 cow dairy in order to determine the effects of vaccination against *Salmonella enterica* serotype Newport on milk production, somatic cell count, and shedding of *Salmonella* organisms (Hermesch *et al*, 2008). Cows were paired (apparently based on their order through the chute system) and one member of a pair of cows was randomly assigned, using a coin toss, to receive *Salmonella* Newport SRP vaccine or control (the vehicle without the antigens) solution. The other cow received the alternate treatment. Vaccine or control solution was injected 45 to 60 days before parturition, and cattle received a second dose 14 to 21 days before parturition. Outcomes included milk production and somatic cell count for the first 90 days of lactation, isolation of *Salmonella* and *Salmonella* Newport antibody levels. Faeces for isolation of *Salmonella* and blood samples for detection of antibodies were collected at the day of first injection and at days 7-14 and 28-35 of lactation).

11.7 FOLLOW-UP/COMPLIANCE

The practical issues involved in managing and conducting a controlled trial have been well described by Knatterud (2002). One important item is to ensure that all groups are followed rigorously and equally (Example 11.5). This is a simpler process if the observation period following the intervention is short, but this time period must be long enough to ensure that all outcomes of interest have been observed and recorded. Regardless of the effort expended on follow-up, it is inevitable that some individuals will be lost to the study through drop-out or lack of compliance. Thus for studies with long follow-up periods, the status of all study subjects should be ascertained at regular intervals throughout the follow-up period (see (Sithole *et al*, 2006)).

A major factor in minimising losses from the study is regular communication with all participants. Incentives to remain in the study might also be provided. These might include financial incentives, provision of information which they might not otherwise have (*eg* detailed udder-health evaluation of a dairy herd provided to participants in a controlled trial of a new dry-cow antibiotic product), or public recognition of their efforts (provided confidentiality concerns have been addressed). For those participants that do drop out, information about study subjects might still be available through routine databases (*eg* milk-production recording programmes) if the participant is willing to provide access. This can be used to either provide some follow-up information or to compare general characteristics of the study subjects withdrawn from the study with those that remained in the study. Nonetheless, because participants in a trial should always have the opportunity to withdraw their animal(s) from a trial, procedures for evaluating those withdrawals should be put in place. This should include methods of documenting the reason for the withdrawal and, potentially, procedures to collect samples from all subjects being withdrawn before their departure. In any event, any losses should be recorded, at specified time points, throughout the conduct of the trial.

In addition to maximising retention in a study, effort needs to be expended to determine if study subjects are complying with the protocol. This might be evaluated through interviews at periodic visits or through collection of samples to test for levels of the drug being investigated. Indirect assessment might be carried out by methods such as collecting all empty containers from products used in a trial. The amount of product (or placebo) used should be appropriate for the number of subjects in the study.

11.8 MEASURING THE OUTCOME

A controlled trial should be limited to 1 or 2 primary outcomes (*eg* disease occurrence in a trial of a prophylactic agent) and a small number (1-3) of secondary outcomes (*eg* productivity, longevity). Having too many outcomes can lead to a problem of 'multiple comparisons' in the analysis (see Section 11.9.1). In addition, if multiple outcomes are measured the intervention may have a different effect on each outcome. Whether or not to combine multiple outcome events into a single composite measure (*eg* a global measure of health by combining scores or occurrences of several diseases) has been the subject of much debate (Ferreira-Gonzalez *et al*, 2007) but for our purposes we prefer designs based on a limited number of primary and secondary hypotheses. When selecting outcomes to be measured, those that can be assessed objectively are preferred to subjective outcomes, but the latter cannot always be avoided (*eg* occurrence of clinical disease). If the outcome is not assessed by a near-gold-standard procedure, the impact of the intervention on the true outcome may differ from the surrogate outcome (Kassai *et al*, 2005). Gilbody *et al* (2007) noted that trials with concurrent economic analyses often contain upwardly biased estimates of intervention effect; thus, if economic analyses are to be included, these should be specified *a priori*.

In general, outcomes should be clinically relevant. Intermediate outcomes, (*eg* antibody titres in a vaccine trial) might be useful in determining why an intervention might not produce the desired outcome, but should not be a replacement for a primary, clinically relevant, outcome related to the objectives of the study (*eg* occurrence of clinical disease). Clinically relevant outcomes include the following:

- diagnosis of a particular disease—requires a clear case definition
- mortality—objective but still requires criteria to determine the cause (if relevant) and time of death
- clinical signs scores for assessing the severity of disease—difficult to develop reliable scales
- objective measures of clinical disease—(*eg* rectal temperature for assessing severity of respiratory disease in feedlot cattle, blood samples to assess the extent of dehydration *etc*)
- measures of subclinical disease—(*eg* somatic cell counts as indicators of subclinical mastitis)
- objective measures of productivity/performance—(*eg* milk production, measures of reproductive performance, weight gain).

Outcomes might be measured on a continuous scale, or as categorical data (often dichotomous), or time-to-event measurements (*eg* time to the occurrence of a disease). Studies based on time-to-event data might have greater power than a study based on simple occurrence, or not, of an event in a defined time period. Outcomes might also be measured at a single point in time, or assessed multiple times for each subject (longitudinal data).

11.9 ANALYSIS

Analysis can be carried out either on an **intent-to-treat** basis or a **per-protocol** basis. In an intent-to-treat analysis, data from all subjects assigned to a specific intervention are included in that intervention regardless of whether or not they completed the study, or whether or not they complied with the protocol. Such an analysis will provide a conservative estimate of the effect of the intervention, as it is recommended to be used, but might reflect the expected response

when the intervention is used in another population with characteristics similar to the study population. In a per-protocol analysis, only subjects which complied and completed the study as outlined in the protocol are included in the analysis. This approach might provide a good measure of response given that the intervention is used as intended but will likely produce a biased estimate of the intervention effect in future use for 2 reasons. First non-compliance is not likely a random event and non-compliers probably are not representative of all participants assigned to that intervention so the estimate of effect may be biased (see Section 12.2). Secondly, there will always be some non-compliance in future use of the intervention so estimating an effect under an assumption of 100% compliance would be unwise.

An analysis usually starts with a baseline comparison of the characteristics of the groups as a check on the adequacy of the randomisation procedures. This should not be based on an assessment of the statistical significance of the difference among groups, but rather an assessment of their comparability. Differences among the groups, even if not statistically significant, should be noted and taken into consideration in the analyses (see below).

The specific procedures for analysing data from controlled trials will not be covered in this chapter but they are discussed in more detail elsewhere in the book. However, a few specific issues will be touched on.

While randomisation is designed to equally distribute potentially confounding factors across the intervention groups, it might not remove all potential **confounding**, especially with small sample sizes (hence the rationale for examining this as noted above). When the outcome is dichotomous, adjustment for covariates is recommended. The best approach is to identify strong predictors a priori, the next best option is to control for covariates that are predictive of the outcome in the trial data (Hernandez *et al*, 2004). Adjusted results should be less biased if the adjustment procedure has removed any residual confounding (particularly a concern in small trials). Adjustment for non-confounders does little harm, provided they are not intervening variables (see Chapter 13). If the outcome is continuous, control of other factors might improve the precision of the estimate of the intervention effect by substantially reducing the unexplained variance.

When measurements are made **before and after** the intervention is administered, it is often useful to adjust for the baseline (pre-intervention) level in each subject when evaluating the response to the intervention. This can either be done by subtracting the pre-intervention value from each post-intervention measurement (*ie* analysing the change in the outcome) or by including the baseline level as a covariate in an analysis of the post-intervention values. Either approach will result in a gain in power for the study, particularly if the correlation between the baseline and the post-intervention measurement is >0.5 (Borm *et al*, 2007b).

Many controlled trials involve repeated assessments of subjects throughout the study period (**longitudinal data**). Analysis of longitudinal data presents some unique challenges. For a starting point the investigator needs to determine if they are most interested in an average effect following intervention, a change in the effect over time or a total effect. Methods of dealing with repeated measures data are covered in Chapter 23. Twisk and de Vente (2008) review methods for dealing with repeated measurements in RCTs. They suggest that if GEE (Chapters 20 and 23)or a similar approach is used for analysis and if the outcome is measured on a continuous scale, only the first follow-up should be adjusted for the baseline (pre-intervention) level of the outcome. Using their approach we would do the following.

- First, perform a linear regression analysis between the first follow-up measurement and

the baseline value.
- Second, calculate the difference between the observed value at the first follow-up measurement and the predicted value from that regression analysis. This difference is called the 'residual change'.
- Third, use this 'residual change' in place of the actual first outcome value in the subsequent GEE analysis.

Longitudinal data often have **missing values** for some of the observations. The problem of missing data is briefly introduced in Section 15.5 and more detailed discussion of the issue can be found in Peduzzi *et al* (2002) and Auleley *et al* (2004). If more than a few observations are missing, the analysis and interpretation will have to take this into account.

Finally, if study subjects are maintained in groups (**clustered data**), it is important to account for the effects of the groups. This is particularly important in cluster randomised trials, but might also be important in trials in which randomisation occurred within the group. Procedures for analysing clustered data are presented in Chapters 20-22.

11.9.1 Multiple comparisons

Controlled trials often give rise to analyses in which **multiple comparisons** are made. There are 3 ways in which multiple comparisons can arise in the analysis of RCTs: examining multiple outcomes, examining multiple subsets of the data, and performing periodic interim analyses during the trial. The problem with multiple comparisons is that the experiment-wise error rate is often much larger than the error rate applied to each single analysis (usually 5%; see Section 15.8.2). This can result in the declaration of spurious effects as significant.

There are many procedures for adjusting the analyses to account for these multiple analyses (Korn & Freidlin, 2008). One of the simplest ways to retain an appropriate experiment-wise error rate is the **Bonferroni adjustment**. This requires that each analysis be carried out using an α/k Type I error rate, where α is the normal error rate (often 0.05) and k is the number of comparisons made. However, this results in a very conservative estimate of the statistical significance of each evaluation. Other, less conservative, procedures can be found in standard statistical texts.

The problem of **subgroup analyses** deserves special attention (Brookes *et al*, 2004). While it is tempting to evaluate a wide range of subgroups within a trial to determine if an intervention had an effect in them, only analyses planned a priori, should be carried out. Otherwise, there is serious danger of identifying spurious associations. Many researchers recommend that findings from unplanned subgroup analyses be reported as exploratory. Furthermore, the recommended approach to ascertain if the intervention effect differs by subgroup is to conduct one overall test of interaction between the intervention and the subgroup identifier. Bear in mind that the sample size of the study usually was based on a single overall test of significance not on a per-subgroup basis and in many instances subgroup analyses will have insufficient power to detect meaningful effects. Brookes *et al* (2004) also describe a method to determine the appropriate sample size required to investigate such interactions reliably. As a guideline, effects sizes of at least twice the magnitude of the assumed overall effect have a similar power of detection to that of the overall intervention effect.

Sequential design studies (also called 'monitored' studies) are those in which, by design, planned periodic analyses of the data are carried out throughout the trial. These analyses are

carried out so the trial can be stopped if there is:

- clear (and statistically significant) evidence of the superiority of one intervention over another
- convincing evidence of harm arising from an intervention (regardless of the statistical significance of that finding)
- little likelihood that the trial will produce evidence of an effect, even if carried to completion. (This concern is not relevant if the goal of a trial is to demonstrate that a new product/procedure has the same efficacy as an existing standard therapy.)

While sequential designs seems like a logical approach, they tend to lack power (on a per-subject basis), and hence their usage should be restricted to those situations where the benefits are clear.

Interim analyses should not be conducted unless the trial is designed to accommodate them. Methods for interim analyses and for adjusting the sample size to accommodate the procedures are beyond the scope of this text but are reviewed in Todd (2007). One example of stopping a trial based on interim analyses was the badger-control trial in England. Interim analyses (the study was not specifically designed for these) revealed that cattle herds in areas where reactive culling was used had increased levels of bovine tuberculosis; hence this arm of the 'badger control trial was halted (Donnelly *et al*, 2003). Bassler *et al* (2008) point out that stopping a trial early because the intervention appears to be having a very positive effect often results in claims of excess efficacy (*ie* the coefficients overestimate the true effect).

11.10 Clinical trial designs for prophylaxis of communicable organisms

The standard designs discussed thus far need to be modified when the intervention is a prophylactic against a communicable organism (*eg* a vaccine or an anthelmintic). Here we will explain why this modification is needed and make some suggestions about trial designs. See Chapter 27 for a discussion of issues related to infectious disease epidemiology.

When estimating the 'protective ability' of a prophylactic against communicable organisms we need to consider whether we are measuring protection at the individual or at the population level. Furthermore, we need to recognise that the protection we observe can be strongly influenced by:

- the baseline level of transmission of the agent in the population of interest,
- the effectiveness of the vaccine (this is of course what we want to estimate), and
- the percentage of the population we chose to vaccinate in our evaluation of the vaccination strategy.

In a population, disease spreads from subject to subject, either directly or via vehicles contaminated with the organism of interest. The rate of transmission depends on the number of adequate contacts a susceptible subject makes with an infected subject or contaminated vehicle per time period (*eg* per day) (See Section 27.3 for a discussion of infectious disease transmission). Given a reasonable limit to the number of contacts each susceptible subject makes per day, if some of these contacts are with vaccinated subjects and if vaccinated individuals are completely or partially protected against infection, the rate of spread of the disease through the population is decreased. In general, the number of adequate contacts each individual makes and the baseline transmission level depend on the characteristics of the study

groups. Consequently, "2 different randomised, double-blind, placebo-controlled studies taking place in sites that differ by the level of transmission would report different estimates of vaccine efficacy even if the level of protection conferred by the vaccine to a specified challenge to infection is the same in both studies" (Struchiner & Halloran, 2007). In addition, in order to understand disease spread, it is helpful to know whether transmission of an agent within subunits of the population (*eg* herds/flocks) is density or frequency dependent (see also Example 27.3). In density dependent transmission, disease transmission is the same among units of different sizes when the proportion of initially infected subjects is the same. In frequency dependent transmission, transmission increases with the number of initially infected individuals (so larger herds would have greater transmission even if they had the same proportion initially infected). Herd size is commonly associated with the frequency of infectious diseases, but it is often not clear whether the effect is density or frequency dependent.

Prophylaxis can have a number of benefits; first, it can prevent infection given exposure. Second, it can prevent clinical disease or reduce the severity of infection among the infected and this can lower the onward transmission of the agent. Whether infection or disease is the chosen endpoint often depends on the context and on the incubation period of the disease: if short, disease is often the endpoint; if long, infection is usually the endpoint. The ability to reduce the severity or duration of disease among those receiving the prophylactic may have a larger impact on the transmission probability in the population than the ability to protect against infection in individuals. The key is that the protective effect of prophylaxis can differ depending on the endpoint evaluated.

As an example, the usual measure of vaccine efficacy (*VE*; for simplicity, we will not differentiate between infection vs disease as outcomes), at the individual level, is typically measured as:

$$VE_d = \frac{(I_{nv} - I_v)}{I_{nv}} \qquad \textbf{\textit{Eq 11.1}}$$

where I_{nv} and I_v are the incidence rates of the outcome in non-vaccinated and vaccinated individuals respectively (Halloran *et al*, 2007). We have added the subscript 'd' to denote that this is the direct efficacy of the vaccine. Of course, to ascertain the true VE_d we would like to compare counterfactuals (see Section 1.7); namely, the incidence of the outcome in the vaccinated subjects contrasted to what the incidence would have been if the subjects were non-vaccinated. Since we cannot observe these events, we estimate the *VE* by randomly assigning half (or some other proportion) of the study subjects to receive the vaccine and half to get a placebo; both vaccinated and non-vaccinated subjects are free to intermingle in the population. Unfortunately, the measure of *VE* we obtain from using this design is likely to be confounded by the proportion of the study population that is vaccinated. We will explain the rationale for this statement subsequently.

Because the direct *VE* measure often is biased and may only be a small proportion of the total efficacy, epidemiologists are more interested in population-based measures of vaccine effectiveness (Carpenter, 2001b). The total effect of prophylaxis is a population measure and consists of 2 components: the direct or individual level vaccine efficacy (VE_d) noted above and the indirect (VE_{ind}) vaccine efficacy. The indirect vaccine efficacy is a population-based measure and is found by comparing the frequency of the outcome in non-vaccinated animals mixed with vaccinated animals from the randomised study area (here designated population A) to the frequency in non-vaccinated animals from a similar population of non-vaccinated animals

(here designated population B) as follows:

$$VE_{ind} = \frac{I_{nvB} - I_{nvA}}{I_{nvB}} \qquad \textit{Eq 11.2}$$

where I_{nvA} and I_{nvB} are the incidence rates (or risks) in populations A and B, respectively. This indirect effect often is a major component of what is referred to as **herd immunity**. The phenomenon of herd immunity provides protection to non-vaccinated susceptible individuals by interfering with transmission of the agent beyond the direct protective effects in vaccinated individuals. For example, in Section 27.4.3, it is shown that if an infected animal typically contacts 5 susceptible animals, a vaccine that is 80% effective will be expected to stop all transmission of the agent (as a result of herd immunity). Since achieving 100% vaccination coverage can be very difficult and knowing that vaccination levels below 100% can be effective in eliminating disease agents, ascertaining the critical level of vaccination that is required to eliminate a specific disease agent (*eg* rabies virus), is a key component of research on population disease control (Longini *et al*, 1998).

The total effect of the prophylaxis (VE_{tot}) is a weighted combination of VE_d and VE_{ind} and can be estimated using:

$$VE_{tot} = \frac{I_B - I_A}{I_B} \qquad \textit{Eq 11.3}$$

Knowing the total effect of a vaccine provides much more useful information in terms of disease control than does the usual direct measure of vaccine efficacy.

11.10.1 Design and analysis issues for estimating vaccine efficacy

A number of different trial designs can be used to obtain estimates of vaccine efficacy. For example, we can employ a cluster randomised trial design in which we compare the disease frequencies in fully vaccinated versus non-vaccinated populations (within a feedlot we might fully vaccinate a number of pens of cattle and contrast the incidence of disease to that in a number of pens of non-vaccinated cattle as in Example 11.6). This could be extended to dairy or swine herds where we could contrast vaccinated versus non-vaccinated herds. Riggs and Koopman (2004) developed a model of transmission with group randomisation, and in 2005, they noted that if cluster randomisation is used, it increases the power of the study if the majority of transmission is from within the cluster, but decreases the power if most transmission comes from outside the cluster. They also note that, when using cluster randomisation, it is advantageous to sample study subjects and determine their natural level of immunity (*ie* prior to vaccination). This allows for the adjustment for natural immunity prior to assessing vaccine induced immunity While this is perhaps the best approach to obtain valid estimates of *VE*, it becomes very expensive and it does not extend easily to the situation where natural stable groupings of study subjects are not available (*eg* conducting vaccine trials in dogs (for influenza), foxes (for rabies) or badgers (for bovine tuberculosis)). Nor does this approach reflect what might happen in populations where it is unlikely that 100% vaccine coverage will be obtained. Furthermore (as noted above), because of the indirect effects of a vaccine, there is a critical level of vaccination, often considerably below 100%, that will protect the population and potentially lead to eradication of the organism. In order to estimate this, we would have to assign different levels of vaccination (say 25%, 50% and 75%) to groups without exceeding the critical fraction vaccinated that would eliminate disease in the non-vaccinated subjects.

As noted, for disease control, the total effectiveness of the vaccine in the population is of more interest than VE_d. A suggested approach is to use a design that will allow the estimation of the direct, indirect and total vaccine efficacies described above. To implement this, we need at least 2 comparable populations of subjects (more than 2 populations would provide much better estimates of vaccine efficacies but these could become prohibitively expensive). We would randomly assign a proportion of individuals in one population, denoted population A, to receive the vaccine and the remainder to receive a placebo. Subjects in the other similar population, denoted population B, would all remain non-vaccinated (Halloran, 2006). Ensuring exchangeability (*ie* that the populations are similar in all important characteristics that affect the outcome) is a difficult task. Perhaps the most important characteristic they should share is the same level of transmission as this greatly affects the indirect efficacy. In addition, it is important that the 2 populations are fully separated from each other so there is no intermixing of subjects. Since the total effectiveness is a population level measure, several populations are needed for statistical evaluation. Nonetheless, this design is feasible under selected circumstances, and this concept is the basis for interpreting population effects of vaccines (Glezen, 2006). Glezen notes that although the direct effects may be small, the impact on population levels of disease can be very marked. If the disease frequency is judged to be stable, information on the level of disease in the non-vaccinated population can be supplemented with data on the level of disease (in population A) prior to the vaccine trial. The building blocks for the calculations are the outcome frequencies in each of 2 populations within one of which there are vaccinated and non-vaccinated subjects, and in the other only non-vaccinated subjects. Example 11.10 shows an example of the computation of the direct, indirect and total effects of a vaccination program in this scenario.

Another alternative design, that can be used when obtaining these 2 ‘similar’ populations is a difficult task but when there is natural clustering of subjects within the population of concern (*eg* of badgers within setts), is to randomly assign vaccination to half the subjects (*ie* badgers) in a geographical area and subsequently investigate the spread of infection/disease within the clusters (*ie* setts) relative to the proportion of individuals in the cluster that were vaccinated (Longini, *et al*, 1998). However, one would need to ensure some stability to the population of these subsets over the duration of the prophylactic trial.

Carpenter (2001a) discusses the estimation of sample size for prophylactic trials. The major concern is that if the indirect effects are large, the overall incidence of disease decreases and the level of the outcome in the non-vaccinated subjects becomes closer to that in the vaccinated subjects and one could conclude that, based on the individual-level measure of vaccine efficacy, the vaccine is not effective when it really is (such studies lack power because of the herd immunity effect). Detailed consideration of power is discussed by Riggs and Koopman (2005) but is beyond the scope of this text.

11.11 ETHICAL CONSIDERATIONS

There are 2 major components to the ethical considerations for controlled trials of animal-health products and procedures. The first is an ethics review by a board whose focus is the ethical treatment of the participants, and the second is a review by an animal-welfare committee whose focus is the well-being of the animal subjects. Specific regulations and guidelines will vary from country to country, but in general the following issues must be considered.

- Is the investigation justifiable (*ie* is it likely to produce meaningful results which will ultimately benefit animal health)?

Example 11.10 Efficacy of dose regimen and observation of herd immunity from a vaccine against *Escherichia coli* O157:H7 for feedlot cattle

A clinical trial was conducted to test the effect of a vaccine containing Type III secreted proteins of *Escherichia coli* O157:H7 on the probability that feedlot steers would shed *E coli* O157:H7 in faeces (Peterson *et al*, 2007). 480 steers (designated Population A) were assigned randomly to 60 pens (8 head per pen) and to 1 of 4 vaccination treatments (120 cattle per treatment, 2 head per treatment per pen). The 4 treatments were (i) no vaccination; (ii) 1 dose (day 42); (iii) 2 doses, (days 0 and 42); and (iv) 3 doses (days 0, 21 and 42). The placebo was the adjuvant and carrier. Another 128 steers (designated Population B) were assigned randomly to 12 pens within the same feedlot to serve as non-vaccinated external controls. The relevant proportions (incidence risks) of cattle shedding E. coli O157:H7 were:

- RnvA = 12% (risk in non-vaccinates in population A)
- RvA = 9% (risk in vaccinates in population A)
- RB = 29% (risk in population B—all non-vaccinates)
- RA = 10.5% (overall risk in population A)

The direct effectiveness of vaccination would be measured by comparing the outcome frequency in vaccinated and non-vaccinated subjects within population A

$$VE_d = \frac{R_{nvA} - R_{vA}}{R_{nvA}} = \frac{12-9}{12} = 0.33$$

In this instance the direct effect of vaccination is to reduce infection in vaccinates by about 33%

The indirect effectiveness would be measured by comparing the outcome frequency in the non-vaccinated subjects in the 2 populations:

$$VE_{ind} = \frac{R_{nvB} - R_{nvA}}{R_B} = \frac{29-12}{29} = 0.59$$

Here there is a large indirect effect of almost 60%

The total effectiveness of vaccination would be measured by comparing the crude outcome frequency in the 2 populations

$$VE_t = \frac{R_B - R_A}{R_B} = \frac{29-10.5}{29} = 0.64$$

Overall, the vaccination protocol reduced the level of infection by approximately 64%, and most of this effect was due to the indirect effects of vaccination. Had we only focused on the direct effects we might have concluded that the vaccine was not very effective.

- Has the design of the study been adequately planned to ensure that valid results will be obtained?
- Is the sample size appropriate? In this case, the need of an adequate sample size to ensure sufficient power for the study will have to be balanced by a desire to minimise the sample size in order to reduce the number of subjects who might receive the less desirable intervention.
- Are procedures in place to minimise the risk and maximise the benefits for participants and subjects in the study? This consideration, and the preceding one might necessitate interim analyses of results.
- Are all participants in the trial enrolled on the basis of informed consent? The provision of informed consent implies that not only have they had the details of the trial provided to

them, but this has been done in a manner that ensures that they understand both the risks and benefits of participating.
- Participants must have the option to withdraw from the study if they so choose.
- Has adequate provision been made to protect all data to ensure their confidentiality, and to ensure that the completeness and accuracy of the data are maintained.

11.12 Reporting of clinical trials

Poor quality of reporting of trials remains a problem (Berwanger *et al*, 2008; Burns & O'Connor, 2008), although the standard of reporting has improved following the release of the CONSORT statement (Kane *et al*, 2007; Moher *et al*, 2005). Recently, the CONSORT statements were modified to better serve the needs of investigators working with livestock, and these are presented in Table 11.1 These reporting standards should serve as guidelines to help ensure that critical issues in study design, implementation and eventual reporting are addressed during the planning of the study. Each of us should be aware of the common biases that can impact the design and reporting of trial results in order to minimise these (Gluud, 2006).

Table 11.1 Checklist of items for the REFLECT-LFS statement: Reporting guidElines For randomisEd Control Trials in Livestock and Food Safety. (For details of Reflect Statement, see www.reflect-statement.org)

Paper section and topic	Item	Modification for trials in livestock species with production, health, and food safety outcomes
TITLE & ABSTRACT	1	How study units were allocated to interventions (eg "random allocation", "randomised", or "randomly assigned"). Clearly state whether outcome was result of natural exposure or of a deliberate agent challenge
INTRODUCTION Background	2	Scientific background and explanation of rationale
METHODS Participants	3	Eligibility criteria for owners/managers and study units at each level of organisational structure, and settings and locations where data were collected
Interventions	4a	Precise details of interventions intended for each group, level at which intervention was allocated, and how and when interventions were actually administered
	4b	Precise details of agent and challenge model, if a challenge study design was used
Objectives	5	Specific objectives and hypotheses. Clearly state primary and secondary objectives (if applicable)
Outcomes	6	Clearly defined primary and secondary outcome measures and level at which they were measured, and, when applicable, any methods used to enhance quality of measurements (eg multiple observations, training of assessors)
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules. Sample size considerations should include sample size determinations at each level of organisational structure and assumptions used to account for any non-independence among groups or individuals within a group
Randomisation—Sequence generation	8	Method used to generate random allocation sequence at relevant level of organisational structure, including details of any restrictions (eg blocking, stratification)
Randomisation—Allocation concealment	9	Method used to implement random allocation sequence at relevant level of organisational structure (eg numbered containers) clarifying whether sequence was concealed until interventions were assigned
Randomisation—Implementation	10	Who generated allocation sequence, who enrolled study units, and who assigned study units to their groups at relevant level of organisational structure
Blinding (masking)	11	Whether or not those administering interventions, caregivers and those assessing outcomes were blinded to group assignment. If done, how success of blinding was evaluated. Provide justification for not using blinding if it was not used.
Statistical methods	12	Statistical methods used to compare groups for all outcome(s). Clearly state level of statistical analysis and methods used to account for organisational structure, where applicable. Methods for additional analyses, such as subgroup analyses and adjusted analyses
RESULTS Study unit flow	13	Flow of study units through each stage for each level of organisation structure of study (a diagram is strongly recommended). Specifically, for each group report numbers of study units randomly assigned, receiving intended treatment, completing study protocol, and analysed for primary outcome. Describe protocol deviations from study as planned, together with reasons
Recruitment	14	Dates defining periods of recruitment and follow-up
Baseline data	15	Baseline demographic and clinical characteristics of each group, explicitly providing information for each relevant level of organisational structure. Data should be reported so that secondary analysis, such as risk assessment, is possible

Paper section and topic	Item	Modification for trials in livestock species with production, health, and food safety outcomes
Numbers analysed	16	Number of study units (denominator) in each group included in each analysis and whether analysis was by "intention to treat". State results in absolute numbers when feasible (eg 10/20, not 50%)
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, accounting for hierarchy, and estimated effect size and its precision (eg 95% CI)
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory
Adverse events	19	All important adverse events or side effects in each intervention group
DISCUSSION Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and dangers associated with multiplicity of analyses and outcomes. Where relevant, a discussion of herd immunity should be included. If applicable, a discussion of relevance of disease challenge should be included.
Generalisability	21	Generalisability (external validity) of trial findings.
Overall evidence	22	General interpretation of results in context of current evidence.

REFERENCES

Auleley GR, Giraudeau B, Baron G, Maillefert JF, Dougados M, Ravaud P. The methods for handling missing data in clinical trials influence sample size requirements J Clin Epidemiol 2004; 57: 447-53.

Barthel FM, Babiker A, Royston P, Parmar MK. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over Stat Med 2006; 25: 2521-42.

Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G. Early stopping of randomized clinical trials for overt efficacy is problematic J Clin Epidemiol 2008; 61: 241-6.

Berge AC, Lindeque P, Moore DA, Sischo WM. A clinical trial evaluating prophylactic and therapeutic antibiotic use on health and performance of preweaned calves J Dairy Sci 2005; 88: 2166-77.

Berwanger O, Ribeiro RA, Finkelsztejn A, Watanabe M, Suzumura EA, Duncan BB, Devereaux PJ, Cook D. The quality of reporting of trial abstracts is suboptimal: Survey of major general medical journals J Clin Epidemiol 2009; 62: 387-92.

Borm GF, Fransen J, Lemmens WA. A simple sample size formula for analysis of covariance in randomized clinical trials J Clin Epidemiol 2007a; 60: 1234-8.

Borm GF, van der Wilt GJ, Kremer JA, Zielhuis GA. A generalized concept of power helped to choose optimal endpoints in clinical trials J Clin Epidemiol 2007b; 60: 375-81.

Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable J Clin Epidemiol 2005; 58: 1220-6.

Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test J Clin Epidemiol 2004; 57: 229-36.

Burns MJ, O'Connor AM. Assessment of methodological quality and sources of variation in the magnitude of vaccine efficacy: a systematic review of studies from 1960 to 2005 reporting immunization with Moraxella bovis vaccines in young cattle Vaccine. 2008; 26: 144-52.

Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials. Stat Med 2007; 26: 2-19.

Campbell PJ, Hammell KL, Dohoo IR, Ritchie G. Historical control clinical trial to assess the effectiveness of teflubenzuron for treating sea lice on Atlantic salmon Dis Aquat Organ 2006a; 70: 109-14.

Campbell PJ, Hammell KL, Dohoo IR, Ritchie G. Randomized clinical trial to investigate the effectiveness of teflubenzuron for treating sea lice on Atlantic salmon Dis Aquat Organ 2006b; 70: 101-8.

Carpenter TE. Use of sample size for estimating efficacy of a vaccine against an infectious disease Am J Vet Res 2001a; 62: 1582-4.

Carpenter TE. Evaluation of effectiveness of a vaccination program against an infectious disease at the population level Am J Vet Res 2001b; 62: 202-5.

D'Agostino RB, Sr., Campbell M, Greenhouse J. Non-inferiority trials: continued advancements in concepts and methodology (special papers for the 25th Anniversary of Statistics in Medicine 25(7)). StatMed. 2006;25(7):1097-9.

Donnelly CA, Woodroffe R, Cox DR, Bourne J, Gettinby G, Le Fevre AM, McInerney JP, Morrison WI. Impact of localized badger culling on tuberculosis incidence in British cattle Nature. 2003; 426: 834-7.

Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials Am J Public Health 2004; 94: 416-22.

Donner A, Taljaard M, Klar N. The merits of breaking the matches: a cautionary tale Stat Med 2007; 26: 2036-51.

Dragalin V, Fedorov V. Design of multi-centre trials with binary response Stat Med 2006; 25: 2701-19.

Fedorov V, Jones B. The design of multicentre trials Stat Methods Med Res 2005; 14: 205-48.

Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, Walter SD, Guyatt GH. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns J Clin Epidemiol 2007; 60: 651-7; discussion 658-62.

Gilbody S, Bower P, Sutton AJ. Randomized trials with concurrent economic evaluations reported unrepresentatively large clinical effect sizes. J Clin Epidemiol. 2007;60(8):781-6.

Gluud LL. Bias in clinical intervention research Am J Epidemiol 2006; 163: 493-501.

Glezen WP. Herd protection against influenza. J Clin Virol 2006; 37: 237-43.

Golub HL. The need for more efficient trial designs Stat Med 2006; 25: 3231-5; discussion 3313-4, 3326-47.

Halloran ME. Overview of vaccine field studies: types of effects and designs J Biopharm Stat. 2006; 16: 415-27.

Halloran ME, Hayden FG, Yang Y, Longini,I M,Jr, Monto AS. Antiviral effects on influenza viral transmission and pathogenicity: observations from household-based trials Am J Epidemiol 2007; 165: 212-21.

Hendrick SH, Kelton DF, Leslie KE, Lissemore KD, Archambault M, Bagg R, Dick P, Duffield TF. Efficacy of monensin sodium for the reduction of fecal shedding of Mycobacterium avium subsp. paratuberculosis in infected dairy cattle Prev Vet Med 2006; 75: 206-20.

Hermesch DR, Thomson DU, Loneragan GH, Renter DR, White BJ. Effects of a commercially available vaccine against Salmonella enterica serotype Newport on milk production, somatic cell count, and shedding of Salmonella organisms in female dairy cattle with no clinical signs of salmonellosis Am J Vet Res 2008; 69: 1229-34.

Hernandez AV, Steyerberg EW, Habbema JD. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements J Clin Epidemiol 2004; 57: 454-60.

Hofmeijer J, Anema PC, van der Tweel I. New algorithm for treatment allocation reduced

selection bias and loss of power in small trials J Clin Epidemiol 2008; 61: 119-24.

Hrobjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding Int J Epidemiol 2007; 36: 654-63.

Jaeger GT, Larsen S, Moe L. Stratification, blinding and placebo effect in a randomized, double blind placebo-controlled clinical trial of gold bead implantation in dogs with hip dysplasia Acta Vet Scand. 2005; 46: 57-68.

Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement J Clin Epidemiol 2007; 60: 241-9.

Kassai B, Shah NR, Leizorovicza A, Cucherat M, Gueyffier F, Boissel JP. The true treatment benefit is unpredictable in clinical trials using surrogate outcome measured with diagnostic tests J Clin Epidemiol 2005; 58: 1042-51.

Knatterud GL. Management and conduct of randomized controlled trials Epidemiol Rev. 2002; 24: 12-25.

Korn EL, Freidlin B. Conditional power calculations for clinical trials with historical controls Stat Med 2006; 25: 2922-31.

Korn EL, Freidlin B. A note on controlling the number of false positives Biometrics. 2008; 64: 227-31.

Lavori P, Kelsey J. Clinical Trials Epidemiol Rev. 2002; 24: 1-90.

Longini,I M,Jr, Sagatelian K, Rida WN, Halloran ME. Optimal vaccine trial design when estimating vaccine efficacy for susceptibility and infectiousness from multiple populations Stat Med 1998; 17: 1121-36.

Moher D, Schulz KF, Altman D, CONSORT Group. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials Explore 2005; 1: 40-5.

O'Connor AM, Sargeant JM, Gardner IA, Dickson JS, Torrence ME, Dewey CE, Dohoo IR, Evans RB, Gray JT, Greiner M, Keefe G, Lefebvre SL, Morley PS, Ramirez A, Sischo W, Smith DR, Snedeker K, Sofos J, Ward MP, Wills R. The REFLECT statement: methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. Prev Vet Med. 2010 Jan 1;93(1):11-8. (Also published in: J Food Prot, Zoonoses Public Health, J Vet Intern Med, J Swine Health Prod) .

Peduzzi P, Henderson W, Hartigan P, Lavori P. Analysis of randomized controlled trials Epidemiol Rev. 2002; 24: 26-38.

Peterson RE, Klopfenstein TJ, Moxley RA, Erickson GE, Hinkley S, Rogan D, Smith DR. Efficacy of dose regimen and observation of herd immunity from a vaccine against Escherichia coli O157:H7 for feedlot cattle J Food Prot. 2007; 70: 2561-7.

Rassnick KM, Rodriguez CO, Khanna C, Rosenberg MP, Kristal O, Chaffin K, Page RL. Results of a phase II clinical trial on the use of ifosfamide for treatment of cats with vaccine-associated sarcomas Am J Vet Res 2006; 67: 517-23.

Riggs T, Koopman JS. Maximizing statistical power in group-randomized vaccine trials Epidemiol Infect 2005; 133: 993-1008.

Riggs TW, Koopman JS. A stochastic model of vaccine trials for endemic infections using group randomization Epidemiol Infect 2004; 132: 927-38.

Ritter MJ, Ellis M, Anderson DB, Curtis SE, Keffaber KK, Killefer J, McKeith FK, Murphy CM, Peterson BA. Effects of multiple concurrent stressors on rectal temperature, blood acid-base status, and longissimus muscle glycolytic potential in market-weight pigs J Anim Sci 2009; 87: 351-62.

Silva LA, Silva CA, Borges JR, Fioravanti MC, Borges GT, Atayde IB. A clinical trial to assess the use of sodium hypochlorite and oxytetracycline on the healing of digital dermatitis lesions in cattle. Can Vet J. 2005;46(4):345-8.

Sithole F, Dohoo I, Leslie K, DesCoteaux L, Godden S, Campbell J, Keefe G, Sanchez J. Effect of eprinomectin pour-on treatment around calving on reproduction parameters in adult dairy cows with limited outdoor exposure Prev Vet Med 2006; 75: 267-79.

Smith DR, Moxley RA, Peterson RE, Klopfenstein TJ, Erickson GE, Clowser SL. A two-dose regimen of a vaccine against *Escherichia coli* O157:H7 type III secreted proteins reduced environmental transmission of the agent in a large-scale commercial beef feedlot clinical trial Foodborne Pathog Dis 2008; 5: 589-98.

Struchiner CJ, Halloran ME. Randomization and baseline transmission in vaccine field trials Epidemiol Infect. 2007; 135: 181-94.

Todd S. A 25-year review of sequential methodology in clinical studies Stat Med 2007; 26: 237-52.

Treweek S, McCormack K, Abalos E, Campbell M, Ramsay C, Zwarenstein M, PRACTIHC Collaboration. The Trial Protocol Tool: The PRACTIHC software tool that supported the writing of protocols for pragmatic randomized controlled trials J Clin Epidemiol 2006; 59: 1127-33.

Turner RM, White IR, Croudace T, PIP Study Group. Analysis of cluster randomized cross-over trial data: a comparison of methods Stat Med 2007; 26: 274-89.

Twisk JW, de Vente W. The analysis of randomised controlled trial data with more than one follow-up measurement. A comparison between different approaches Eur J Epidemiol 2008; 23: 655-60.

Ural K, Ulutas B, Kar S. Eprinomectin treatment of psoroptic mange in hunter/jumper and dressage horses: a prospective, randomized, double-blinded, placebo-controlled clinical trial Vet Parasitol. 2008; 156: 353-7.

Zou GY, Donner A, Klar N. Group sequential methods for cluster randomization trials with binary outcomes. Clin Trials. 2005;2(6):479-87.

Zwarenstein M, Oxman A, Pragmatic Trials in Health Care Systems (PRACTIHC). Why are so few randomized trials useful, and what can we do about it? J Clin Epidemiol 2006; 59: 1125-6.