

©1987 S. Wayne Martin, Alan H. Meek, Preben Willeberg

Veterinary Epidemiology: Principles and Methods

Part 2: Studying Disease in Animal Populations

Chapter 7: Design of Field Trials

Originally published 1987 by Iowa State University Press / Ames

Rights for this work have been reverted to the authors by the original publisher. The authors have chosen to license this work as follows:

License Information:

1. The collection is covered by the following Creative Commons License:



Attribution-NonCommercial-NoDerivs 4.0 International license

You are free to copy, distribute, and display this work under the following conditions:

	Attribution: You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work.) Specifically, you must state that the work was originally published in <i>Veterinary Epidemiology: Principles and Methods (1987)</i> , authored by S. Wayne Martin, Alan Meek, and Preben Willeberg.
	Noncommercial. You may not use this work for commercial purposes.
	No Derivative Works. You may not alter, transform, or build upon this work.

For any reuse or distribution, you must make clear to others the license terms of this work.

Any of these conditions can be waived if you get permission from the copyright holder.

Nothing in this license impairs or restricts the author's moral rights.

The above is a summary of the full license, which is available at the following URL:

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

2. The authors allow non-commercial distribution of translated and reformatted versions with attribution without additional permission.

Full text of this book is made available by Virginia Tech Libraries at: <http://hdl.handle.net/10919/72274>

Design of Field Trials

This chapter describes how to design and conduct prophylactic and therapeutic field trials. Acquiring such knowledge is particularly relevant today because there has recently been an increasing awareness of the need for such trials. Knowledge of the principles of field trial design allows veterinarians to better interpret the literature in applied as well as research journals. Also, many veterinarians are asked to collaborate in field trials with universities, drug companies, or government drug control and evaluation agencies. In addition, many practitioners utilize field trials to evaluate and improve the disease control regimes they offer their clients.

The essence of the experimental method is the planned comparison of the outcome in groups receiving different levels of a treatment. As an example, the treatment levels might be vaccine versus no vaccine, and the outcomes might be the rate of subsequent disease and the productivity of animals in each treatment group. In the context of this chapter, a treatment could be a therapeutic drug, a prophylactic biologic (e.g., a vaccine), or an entire program composed of many individual treatments (e.g., a preconditioning program for beef calves consisting of weaning, creep feeding, and vaccination). When designing a field trial, one attempts to ensure the comparability of the units receiving each level of the treatment and to reduce the experimental error so that practical treatment effects can be identified with the minimum number of experimental units. The comparability of the treatment groups depends chiefly on the method of allocating the experimental units to treatment groups and the management of the groups during the course of the trial. At the termination of the study, statistical tests are used to evaluate the likelihood that chance variation produced any observed differences in the outcome between treatment groups.

In a manipulative laboratory experiment, the investigator can control the allocation of experimental units to treatment groups and also the timing and nature of the challenge to treatment. In field experiments, the investi-

gator can control the allocation of experimental units but usually has to depend on natural challenge of the treatment. However, careful selection of the experimental units can increase the probability that a sufficient challenge will occur. For ethical reasons, the only field experiments often possible (except for those conducted under artificial conditions such as at research stations) involve treatments having a high probability of being found valuable in preventing or treating the disease(s) of concern.

Frequently, field trials are used to test a specific hypothesis, but in addition they may be used to validate the findings of observational studies or laboratory experiments. Sometimes the results of field trials may provide an indirect evaluation of a causal hypothesis. For example, serologic evidence may incriminate an agent such as the bovine virus diarrhea (BVD) virus as a cause of respiratory disease in feedlot calves. Although it has not been possible to produce respiratory disease with BVD virus alone or in combination with other agents, ancillary findings from experiments and some observational studies give support to the hypothesis that BVD virus infection is a determinant of respiratory disease of feedlot calves. If a properly designed field trial of a BVD vaccine produced a decrease in the occurrence of respiratory disease, this evidence would indirectly support the hypothesis as well as provide a method of control of respiratory disease. If the vaccine did not produce a significant benefit, the hypothesis would not be rejected because the vaccination regime may have been ineffective. Finally, as will be discussed later, trials may be performed to estimate parameters for building computer models.

The important biometric features of field trial design are discussed elsewhere (Cochran and Cox 1957; Snedecor and Cochran 1980; Armitage 1971). In choosing a particular design, the investigator attempts to ensure the field trial results will be valid, the probability of type I and/or II errors is reduced, and the design is practical for the specific field conditions that exist. The application of these features to the design of experiments conducted on humans and/or on privately owned animals has been discussed from a number of viewpoints (Peto et al. 1976, 1977; Gilbert 1974; Byar et al. 1976; Martin 1978). An excellent introduction to clinical trials is provided by Colton (1974).

For a number of reasons, field trials have not been widely or well used in veterinary medicine, at least in terms of assessing the efficacy of vaccines against bovine respiratory disease (Martin 1983). Sir Austin Bradford Hill, a pioneer in the use of field trials in human medicine, concisely summarizes the need for clinical trials. Although directed toward medical doctors, the reader can extrapolate the following statements to therapeutic and prophylactic trials in veterinary medicine (Hill 1952):

Therapeutics is the branch of medicine that, by its very nature, should be experimental. For if we take a patient afflicted with a malady, and we alter his conditions

of life, either by dieting him, or by putting him to bed, or by administering to him a drug, or by performing on him an operation, we are performing an experiment. And if we are scientifically minded we should record the results. Before concluding that the change for better or for worse in the patient is due to the specific treatment employed, we must ascertain whether the results can be repeated a significant number of times in similar patients, whether the result was merely due to the natural history of the disease or in other words to the lapse of time, or whether it was due to some other factor which was necessarily associated with the therapeutic measure in question. And if, as a result of these procedures, we learn that the therapeutic measure employed produces a significant, though not very pronounced, improvement, we would experiment with the method, altering dosage or other detail to see if it can be improved. This would seem the procedure to be expected of men with six years of scientific training behind them. But it has not been followed. Had it been done we should have gained a fairly precise knowledge of the place of individual methods of therapy in disease, and our efficiency as doctors would have been enormously enhanced.

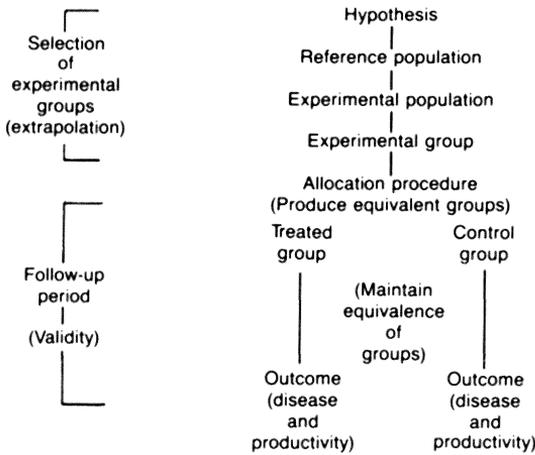
Ethical considerations are an important feature of field trials also, and often influence whether an experiment can be performed. In this regard, Hill goes on to state:

In addition to asking whether it is ethical in the light of current knowledge to plan a randomized trial in which some . . . will not be offered the new measure, it is also necessary to ask whether it is ethical not to plan a randomized trial, since failure to do so may subject the population as a whole to the perpetuation of an ineffective program.

The key items to be considered in the design and performance of a field trial are shown in Figure 7.1. The process of selecting the experimental group is of great importance because it may limit the generalization of the experimental results (i.e., the ability to extrapolate the results beyond the experimental units actually used in the trial). The follow-up period, which extends from the time of allocation to the end of the trial, can greatly influence the validity of the experimental results (i.e., the degree of certainty that the observed results are attributable to the treatment given). If one is forced to choose between validity and the ability to generalize, those issues concerned with validity should receive priority. The remainder of this chapter describes the major features of field trial design and performance in more detail.

7.1 Objective of the Experiment

The objectives should be clearly stated, and both major and minor objectives identified when appropriate. This description should identify the outcome (response variable) and allow the straightforward development of the required treatment contrasts; these are obvious with only two levels of treatment (e.g., new versus standard treatment) but less obvious if three or



7.1. Key items in design and performance of field trial.

more treatment levels are present. In most studies the number of objectives should be restricted, perhaps to one or two; otherwise the design may become very complex and the performance of the trial jeopardized.

7.2 Reference and Experimental Populations

The population that will benefit if the treatment is effective is the reference population. In a prophylactic trial, the reference population consists of “healthy” individuals that are at risk of the disease, whereas in clinical trials it is those with a defined disease or syndrome. In both types of trial the experimental units are allocated to either the new treatment, the standard treatment, or the no treatment group. With due allowance for practical matters of convenience and cost, the population in which the trial is conducted (the experimental population) should be representative of the reference population. This allows the investigator to extrapolate the results of the trial to the larger reference population.

The collaborators in almost all field trials are volunteers. This may lead to concern about the validity of field trials since volunteers are known to differ in many respects from nonvolunteers. However, this fact should not invalidate the results of the trial provided the volunteers are allocated in a formal random manner to the treatment groups. When possible, it is useful to compare the characteristics of the volunteers to those of nonvolunteers as this information is valuable in guiding decisions concerning the

extrapolation of results. Given that the treatment program is found to be beneficial, this information may prove useful in modifying the program to make it more acceptable when it is subsequently offered to members of the reference population.

7.3 Experimental Unit

The experimental unit is the smallest independent grouping of elements (i.e., individuals) that could receive a different treatment given the method of allocation; that is, providing the units are independent, the experimental unit is the smallest aggregate of individuals that is randomized to the treatment groups. Failure to identify the proper experimental unit is common and has serious consequences in terms of interpreting the results of an experiment. Suppose a new treatment was assigned to all animals in one herd with the standard treatment allocated to all animals in a different herd, using a formal random method such as a coin toss. Since a herd is the smallest grouping of animals allocated, such a scheme provides only one experimental unit per group; the number of animals per herd is of little importance. Since there is only one experimental unit per treatment, it is not possible to estimate the within-treatment variability (variance) and no formal statistical evaluation of observed differences is possible. Hence, the results can not be analyzed to establish the probability that chance variation produced the observed differences. Another example of the same mistake occurs when individuals are allocated to a certain treatment group and then housed together. Here the members of the same pen are not independent because extraneous factors (e.g., poor ventilation, infections) would tend to affect the entire pen and could produce a large difference in the outcome between treatment levels. In this situation, one could not separate a treatment effect from a pen effect. Thus the functional experimental unit is the pen. A similar mistake may occur when individuals are randomly assigned to a treatment level, and each individual is tested a number of times throughout the study. For purposes of statistically testing differences in outcome between treatments, individuals are the experimental unit, not the number of tests or samples.

In some experiments with more than one treatment, it may be desirable to have different experimental units in the same field trial. For example, the treatments of lesser importance may be assigned to herds or aggregates of animals, while the treatments of greater importance are assigned to individuals within the herd or group. These are called split-plot designs as described in 7.4.

7.3.1 Criteria for Entry to the Trial

The criteria that a unit must possess to enter the trial should be stated clearly. For example, only farms known to be infected with K99 *E. coli*

would be included in vaccine trials against this organism. In clinical trials it is very important to specify the criteria used to diagnose the condition of interest (e.g., in a trial of alternative treatments of renal failure, it is essential to specify what constitutes renal failure). If these criteria are not valid or are not followed, the outcome can be severely biased. Adequately specifying the criteria for entry is particularly important when units with certain characteristics are to be excluded. For example, herds where pooled colostrum is fed would be excluded from a trial of an *E. coli* bacterin against neonatal diarrhea if individuals within the herd were to be allocated to vaccine or nonvaccine groups. If many investigators are involved, great care should be exercised to ensure the stated criteria for entry are understood and are followed.

7.3.2 Number of Experimental Units

Unless the study is designed as a sequential trial, the approximate number of units to be included in the trial should be determined before the study begins. In sequential trials, the number of units that eventually enter the trial depends on the results obtained during the trial. Sequential designs allow for frequent testing for significant differences between treatment groups so the trial can be stopped as soon as one treatment is found to be superior. These designs also have a maximum allowable sample size, allowing the trial to end if it becomes obvious the effects of the treatments do not differ by any practical amount. Sequential trials are of greatest value when the information about the response in one unit is available before the next unit enters the study (Armitage 1971).

In the field trials discussed here it is assumed that only treatments producing a sufficiently large true treatment effect to make them of practical importance are of interest; that is, the treatment effect must be sufficiently large to be of biologic and/or economic importance to members of the reference population. In a field trial, the observed treatment effect (the simple difference between the outcome in the treatment and control groups) is used to estimate the true treatment effect (the effect that would become known only after completing an infinite number of field trials). In making this inductive inference about the existence of a true treatment effect from the experiment results, there are two possible types of errors, usually designated as type I and type II. A type I error occurs when it is declared on the basis of the trial results that there is a true treatment effect when in fact there is not. A type II error occurs when it is declared on the basis of the trial results that no true treatment effect exists when in fact the treatment produces a worthwhile effect.

If the probability of a type II error (expressed as a proportion) is subtracted from 1, the result is referred to as the "power" of the experiment; the power being the likelihood that the trial will identify a true treatment effect correctly. Similarly, 1 minus the type I error is the confi-

dence level. If the type I error is 5% (0.05) the confidence level is 95% (0.95). This latter value is the probability of the trial resulting in no significant difference when no real treatment effect exists.

Traditionally, there has been more concern about type I than type II errors, and the probability of committing this error is set by convention at or below the 5% level. In the absence of any knowledge about the relative seriousness of type I and II errors, the probability of committing a type II error is frequently set at four times that of the type I error (i.e., < 20%). However, when feasible, the magnitude of the two error rates should be based on the estimated costs—in biologic, humane and economic terms—of committing these errors.

The reason for discussing these error rates is that it is necessary to consider them when estimating the number of experimental units required for the trial. Too many units is expensive and may in a superficial analysis lead to declaring trivial treatment effects as significant. Too few units reduce the power of the trial (i.e., decreases the chance of detecting biologically significant effects).

Formulas appropriate for calculating the required number of experimental units are shown in Table 2.9. Tables of sample sizes are also available in standard statistical texts (Fleiss 1973). Usually it is best to view these determinations as ball park estimations rather than exact requirements. If the number of individuals required for the trial has been estimated using formulas appropriate for the random allocation of individuals, the total number of animals required when aggregates are randomized may need to be increased 4–5 times. For further discussion on this topic see Comstock (1978) and Cornfield (1978).

Although practitioners are encouraged to perform field trials, it should be recognized that it is much better to have one or two large scale investigations than numerous small ones. The reasons are twofold. First, in a trial with an insufficient number of units to give realistic power, the most frequent conclusion is “the differences between treatment groups were not significant”; thus, the trial produces no useful results. Second, if numerous small studies are conducted by different investigators, the probability that a significant difference will be observed in at least one study when no practical true treatment effect exists is considerably greater than 5%. One also needs to be aware of this problem when reviewing the literature because of the bias (called the publication bias) of many investigators to report differences found to be significant, but not to report “negative findings.”

7.4 Allocation of Experimental Units to Treatment Groups

The use of a formal randomization procedure is desirable for the allocation of experimental units. Indeed, it is the use of formal random allocation that provides the primary advantage of field experiments over prospec-

tive cohort studies. In cohort studies the effects of known extraneous factors can be controlled by analysis, matching, or exclusion. Nonetheless, often it is not possible to account for the effects of all known extraneous factors, and the effects of unknown factors must be ignored. In experiments, random assignment is used to protect against any systematic differences in the treatment and control groups. In this way it prevents bias, balances any confounding variables (those factors related to the outcome and treatment), and guarantees the validity of the statistical test. The latter guarantee rests not on the assurance that the groups will be exactly the same with respect to known and unknown factors, but rather that the probability distribution of all possible outcomes of allocation is available. This allows the calculation of the probability (significance level) that differences in the outcome of interest equal to or more extreme than those observed might have arisen solely from the allocation procedure; these are viewed as chance differences.

Randomization can be achieved by flipping a coin, drawing numbers from a hat, through use of random number tables, or by using random number generators such as are available on some calculators and on most computers. Systematic allocation using a random starting point is an acceptable method of randomization under field conditions, but is less preferable than true randomization. To allocate 50 cattle to each of a vaccine and a control group, one could choose 50 random numbers between 1 and 100. The resulting numbers could relate to a tag number or to the sequence the animals follow in passing through a chute facility and would identify the animals to receive the vaccine. Vaccinating the first 50 animals caught and leaving the remainder as controls would not be an acceptable allocation procedure.

Often in a clinical setting, volunteers are sought and then the animals or units belonging to the volunteers are allocated to the treatment groups if they meet the criteria for treatment. It may be preferable to initially seek animals or units that meet the criteria for entry to the trial and then ask the owner to collaborate. Animals or units belonging to owners who refuse are given the treatment the investigator thinks is best; whereas those belonging to volunteers are allocated to the standard or new treatment group on a formal random basis. Subsequently, the outcome in all three groups (volunteer treatment, volunteer control, and nonvolunteer control group) is identified and used in the analysis. This procedure has advantages over initially seeking volunteers, but the investigator needs to take care lest "special clients" are discouraged from agreeing to enter the trial.

7.4.1 Allocation Methods

There are a number of different ways of allocating experimental units to treatment groups, each method being a different experimental design. In selecting a method one basically takes into account: the number and ar-

range of treatments (whether an experimental unit can receive more than one treatment) including practical constraints in the delivery of a treatment (e.g., certain treatments can only be given to an aggregate of individuals, such as litters or pens); monetary constraints that tend to place an upper limit on the total number of experimental units; and maximizing the precision of the experiment (reducing the experimental error) given the previous limitations.

If no important predictors (covariates) of the outcome are identifiable, the units should be individually, randomly allocated. This is the simplest and most frequently used design and is called a completely randomized design. In large field trials ($n > 100$) with relatively homogeneous experimental units, a completely randomized design together with analytic control of potential confounding variables (covariates) is preferred. (Recall that randomization does not guarantee that all covariates will be equally distributed in the treatment groups.) The set of potential confounders must be listed beforehand so that the presence or level of those variables can be noted at the appropriate time during the field trial.

In some field trials, the variability within treatment groups can be decreased by grouping (blocking or matching) the units so they are similar with regard to important characteristics. The units within these blocks are then randomly allocated to treatment groups on a within-block basis; this constitutes a randomized block design. Each block constitutes a replication of the treatments. In clinical trials it may be advisable to "block" on one or two very important prognostic factors (e.g., severity or chronicity of disease, age of patient), ensuring equal distribution of these factors within the treatment groups. The determination of treatment effect is then done on a within-block (pair) basis. Whether to use blocking requires knowledge of the amount of precision gained by matching relative to the loss sustained by the reduction in the number of replications.

Sometimes the experimental unit may serve as its own control; these are called cross-over designs. In cross-over designs, the treatments are allocated to the same unit in random order over a series of periods. If treatments are likely to have a residual effect, the magnitude of the effect should be identified and accounted for prior to testing the significance of any treatment effect. Another strategy is to allow an adjustment period between treatments.

Factorial designs can be used if there are two or more treatments and each experimental unit can receive both treatments. With two different treatments, some units receive no treatment, others one treatment, and still others both treatments. There are two major advantages to factorial designs relative to the traditional method of studying only one factor at a time. First, two treatments can be studied with the same number of units required to assess one treatment. Second, the effects of combining the treat-

ments (additive, synergistic, or antagonistic) can be evaluated. This latter feature is quite important in that many biologics (including vaccines) are given in combination, and their combined effects may differ from their singular effects. To keep the requirements for the trial practical under field conditions as well as to aid interpretation of results, most field trials should have a maximum of three factors.

Split-plot designs are a subtype of factorial designs; the difference is that the experimental unit for one factor (treatment) is different from that for the other factor (treatment). Often this design is chosen when one treatment can only be given to aggregates of individuals (e.g., antimicrobials in the water supply of a litter of pigs or a pen of cattle), whereas the other treatment can be allocated to individuals within the aggregate (e.g., assigning individual pigs or cattle to receive a vaccine). Split-plot designs have the same general advantages of factorial designs, particularly the ability to assess interaction between the treatments. If the whole-plot factor is not randomly assigned (e.g., perhaps the owner decided which litters would receive antimicrobials in the water), one cannot assess the effect of this treatment. However, any interaction between the two treatments can still be assessed. In other instances the split-plot design allows the actual field procedures to be performed with less hassle than if an ordinary factorial design were used (Martin et al. 1984). These five designs (completely randomized, randomized complete blocks, cross-over, factorial, and split-plot) are probably the most common designs used in field trials in veterinary medicine. However, there are a large number of other designs, and the advice of a statistician should be sought early in the planning phase of any field trial.

To avoid severe imbalance in the number of units receiving each treatment at any time throughout the study period, the randomization strategy should allow for equalizing the number in each group at fixed intervals (e.g., after every fourth or eighth unit has entered the trial). This procedure is called "balancing," and prevents temporal factors from biasing the outcome (e.g., severe weather at a time when the new treatment had been used most frequently).

7.4.2 Nonrandom Allocation Methods

Recently there has been much discussion about the ethics of randomized trials, and a number of articles describing alternatives to random allocation have appeared. Most of these methods use prior knowledge of treatment efficacy in addition to the ongoing results of the trial to decide the treatment given to the next experimental unit to enter the trial. Thus these techniques are restricted to trials where the units enter the study over an extended period of time, and the response of one unit can be assessed before the next unit is entered. Many clinical trials may fit this design.

One simple example of these designs known as adaptive allocation is called "play-the-winner." To utilize this method, at the start of the study a consensus is obtained about the treatment to be given to the first unit. If the response to this treatment is favorable, the same treatment is given to the second unit and each following unit until an unfavorable response occurs. When a failure is observed, the alternative treatment is used and continued until a treatment failure occurs, at which point the first treatment is used on the next unit to enter the trial. The process is repeated until the desired number of units has entered the study. This design ensures that the most efficacious treatment is given to the majority of animals in the trial, and this may reduce owner concerns related to the ethics of randomization. However, as the true difference between treatment effects decreases, the advantage of this method over the traditional random allocation is also reduced. Also, if adaptive allocation is used, it is extremely important to define what constitutes a failure. Otherwise, if the identity of the treatment is known, subjective bias in assessing the results may occur.

It should be noted that the use of historical controls (the before and after comparison) has virtually no place in field trials in veterinary medicine. The unpredictability of the outcome and the numerous possible differences between the before and after periods prevent the valid use of this design except in rare circumstances. This technique is particularly prone to bias if herds or flocks with a history of severe disease problems are given a treatment and the current disease status compared to the previous status. In addition to a host of possible differences between the periods that could influence the outcome (e.g., weather) the probability of problem herds getting worse is quite small, the probability of getting better rather large. Hence, almost any treatment may falsely appear to be effective (Acres and Radostits 1976).

7.4.3 Assigning Unequal Numbers of Units to Each Treatment

Assuming that only two treatments are being compared (in a completely randomized design) and in the absence of clear indications about the efficacy of the new treatment, the allocation of equal numbers to the treatment and control groups is preferable. But if evidence exists that the new treatment is likely to be better than the standard treatment, unequal allocation (not to exceed 2:1) can maximize the benefit to those in the trial. That is, 2 experimental units are assigned to the treatment group for each 1 experimental unit allocated to the control group. There is no value in proceeding beyond the 2:1 ratio, because in order to maintain the power of the field trial the total number of units may have to be increased to compensate for the unequal allocation.

7.4.4 Biologic Factors That May Affect Allocation

Certain factors have been described that may necessitate modifications in the design of vaccine and/or therapeutic trials. These factors usually involve an aspect of herd immunity, which allows groups of animals to stop or slow infection and/or minimize its effects, and may exist even when not all individuals within the herd are resistant. In this event, the vaccinated or treated majority may protect or otherwise reduce the challenge to the non-treated minority, minimizing differences in outcome between the treatment groups and leading to the conclusion the treatment was not effective. Also, if the treatment is applied to only a small proportion of the herd, the untreated animals present an unduly large source of infection or challenge, and the study again is biased toward accepting the null hypothesis of no treatment effect (Thurber et al. 1977). This can also happen in testing anthelmintics, since the untreated animals may seed the environment and increase the challenge to treated animals in the same area. The best way of circumventing these problems is to use experimental units that are or can be separated physically from each other (e.g., randomize herds rather than animals within a herd to treatment groups).

Another means of avoiding this problem is to use the herd as its own control in a cross-over design. In doing this, each herd is treated (or not treated) for a specified period of time, and at the end of each time period a decision is made (using a formal random process) whether to treat for the next period. A difficulty with this design is that the residual effects of spread of vaccine organisms and/or herd immunity may extend into adjacent treatment periods. Thus the duration of these periods would have to be carefully defined and/or the residual effects quantified and removed analytically. A further drawback to the cross-over design is the difficulty in ensuring equality of handling of each group if the treatments given in a period are known to the owner and/or the investigator.

7.5 Treatment Regimes

The different treatments (including their timing, method, and route of administration) and dosages should be clearly and completely specified. Besides providing clarity of purpose and performance to the study, it is the total program that is being evaluated, not just a specific treatment. If appropriate, the other treatments or manipulations that can or can not be given should be specified. Otherwise if the cointerventions are related to the treatment, the outcomes may be biased. In most field studies the program given to the comparison (reference) group must be the best treatment currently available. Only when no satisfactory treatment exists can field trials

offering no treatment to the comparison group be justified. When a new treatment is being investigated, the highest recommended and safe dose should be used to increase the validity of negative findings. Where possible, more than one dosage level should be included so that any dose-response relationship can be identified.

7.6 Follow-Up Period

Management or other biases related to both conscious and unconscious beliefs about the value of the various interventions under study may give rise to differences developing between treatment groups. Most frequently, bias will be evidenced in the differential management of or assessment of outcome in members of the treatment and control groups. The simplest and most effective way of minimizing this bias is to prevent knowledge of the treatment status of individual experimental units by using blind techniques and placebos. Blind techniques may be used to keep one party (the owner) or the other (the investigator) or both ("double blind") unaware of the treatment status of any given experimental unit. In this manner, systematic differences between groups in the management of the animals or in assessing the outcome will be minimized.

To maintain blindness it frequently is necessary to use dummy treatments or placebos. Depending upon the situation, the placebo might be an innocuous look-alike antibiotic, a fake vaccine, or any substance or regime designed to mimic the real treatment. If two very different appearing drugs are to be compared, the syringe can be filled and the barrel taped to hide the identity of the drug. In some instances, no amount of camouflage can hide the identity of the treatment regime; nonetheless, it is essential to maintain as much similarity in the management and assessment of both groups as possible.

The effects of other problems such as noncompliance and withdrawal from the trial tend to be reduced through the use of blind techniques. When possible, the extent of compliance should be noted and reasons for withdrawal recorded.

7.7 Measuring Outcome

The outcome or response should be of practical importance to the animal and/or its owner (Burns 1963). Thus, titer response, blood level of drug, or parasite egg count per gram of feces should not be used as substitutes for measuring protection against disease or decreased production, as one may not predict the other (e.g., titer response often is a poor predictor of protection against disease).

For trials performed in domestic animals, at least two outcomes (one

concerned with productivity and the other with morbidity or mortality) should be used, because it is possible for a treatment to affect one outcome but not the other. A treatment may lower morbidity rates but have little effect on growth rate or feed efficiency. This appears to be the case with vaccination against atrophic rhinitis in swine.

In choosing a parameter to represent the outcome or response, preference should be given to those that can be measured objectively and quantitatively over those that must be measured on a subjective basis. However with a little ingenuity, scoring systems can be developed to help increase the precision involved in subjective assessments. In either event it is preferable to measure the outcome without knowledge of the treatment status by using some form of blind technique. Obviously this is most important when the outcome is judged subjectively.

7.8 Analyzing Treatment Effect

The actual effect of the treatment regime is found by comparing the outcome in those members of the treatment and control groups who complied fully with the regime. The effect anticipated if the same treatment were offered in the same manner to similar groups can be obtained by using the original group allocation when calculating the treatment effects. The latter evaluation includes any effects resulting from noncompliance, deviation from original treatment allocation, evaluation, etc.

It is not the intention here to discuss the statistical tests that can be used to analyze the results of field trials. Simple statistical tests such as the student's t-test or the chi-square test will suffice in trials with only two treatment groups. If matching is used, the test should be chosen accordingly. Suffice it to say that the design of the study dictates which statistical test to use. The reader is encouraged to consult an appropriate statistical text for details.

Finally a brief comment should be made on significance levels. Often the statement "significant at the 5% level" is taken as proof of a treatment effect. In fact, the above statement merely indicates that if the only cause of differences in the frequency or extent of outcome is the allocation of experimental units, the likelihood of differences equal to or larger than those observed is less than 5%. Other factors besides the treatment could have produced the differences. In addition, the actual meaning of the significance level is difficult to interpret, because the literature is biased by the tendency to report positive and withhold negative results. For this and other reasons (such as frequent snooping at the data for significant differences), one should be somewhat conservative in interpreting the level of significance reported in the literature.

After reading this chapter, there will still be individuals who believe

that any and all experimentation on clients' animals is unethical or that field trials are too difficult to control. For these individuals the results of observational studies and, more frequently, experience or expert advice form the basis of treatment selection. Unfortunately experience is often a poor method for determining the truth about treatment efficacy. Hence, the randomized trial remains the best current method of assessing treatment efficacy to ensure that we do more good than harm for our clients and their animals.

7.9 Examples of Field Trials

Table 7.1 summarizes some of the data resulting from a field trial of an *E. coli* bacterin and a reo-like virus vaccine in beef cows and calves (Acres and Radostits 1976). This well-designed trial utilized a factorial design to examine the separate and joint effects of the bacterin and virus vaccine. Great effort was taken to obtain valid data, and a placebo was used to minimize management bias in the performance of the trial as well as to prevent potential bias in assessing the outcome. The various rates and calculations are explained well and the discussion section should prove informative to those contemplating field trials. Unfortunately, neither the bacterin nor the virus vaccine appeared to be effective in reducing morbidity or mortality. (The reader may verify this by applying the chi-square test to the data in Table 7.1. The Mantel-Haenszel method of analysis may also be used; for example, to test the calf vaccine effect controlling for the effect of the bacterin given to the cow.)

Table 7.2 summarizes the results of another field trial involving gonadotrophin-releasing hormone (GnRH) (given at day 15 postpartum) and prostaglandins (given at day 24 postpartum) in a 300-cow dairy herd (Etherington 1983). (That only one herd was used is in retrospect the only major drawback of note to the design of this trial.) Again, a factorial design was used and the occurrence of a number of diseases as well as an

Table 7.1. Morbidity and mortality rates in beef calves, classified by vaccine-bacterin status

	Vaccine ^a to calf	Status of calf ^c		Total calves	Morbidity rate (%)	Mortality rate (%)
		Diarrhea	Died			
Bacterin [*]	Yes	32	3	163	19.6	1.8
	No	41	7	172	23.8	4.1
Placebo	Yes	33	10	160	20.6	6.2
	No	34	3	187	19.2	1.6

Source: Acres and Radostits 1976, with permission.

^aK99 *E. coli* bacterin.

^{*}Reolike virus vaccine (now called Rotavirus).

^cWithin 30 days of birth.

Table 7.2. Reproductive performance parameters of 305 Holstein-Friesian cows classified by treatment groups

Outcome parameter	Treatment groups			
	(1) Placebo placebo <i>n</i> = 79	(2) GnRH* placebo <i>n</i> = 73	(3) Placebo prostaglandin <i>n</i> = 76	(4) GnRH prostaglandin <i>n</i> = 77
Days from calving to first observed estrus	60.0	86.4	55.4	61.3
Days from calving to first service	83.9	108.1	91.5	84.5
Days from calving to conception	121.2	135.9	109.9	116.5
Services per conception	1.79	1.65	1.74	1.80
Number of heats detected before first service	0.61	0.5	0.65	0.65
Percent culled for reproductive reasons	10.1	6.8	3.9	3.9
Percent developing pyometra	6.3	17.8	2.6	9.0

Source: Etherington 1983, with permission.

*Gonadotrophin releasing hormone.

important productivity measure (the calving-to-conception interval) were selected as outcomes. Plasma progesterone levels were measured at three different times (days 15, 24, and 28) postpartum to provide an additional and objective biologic indication of the treatment effects. Placebos were used to maintain double-blindness, preventing both differential management and biased assessment of results (e.g., rectal findings). Wherever possible, subjective findings (such as rectal examination results) were quantified (e.g., the actual size of the uterine horn or ovarian follicle was estimated rather than being reported as small, normal, or large). The results of this trial failed to indicate any practical beneficial effect of GnRH on reproductive parameters; in fact the drug appeared to produce adverse effects on the ability of treated cows to conceive, primarily due to an increased occurrence of pyometra. Prostaglandins appeared to produce a beneficial effect alone and also by counteracting the adverse effects of GnRH. The actual analysis used (factorial analysis of variance) is beyond the level of this text; however, the above results should be reasonably apparent after perusing the raw data in Table 7.2. Given that only one herd was used in this study, it makes it difficult to extrapolate results to all dairy cattle. Nonetheless, it points out that what appear to be biologically sensible interventions may not always produce the desired effects.

Neither of the above experiments should be interpreted as providing conclusive evidence about the efficacy of the biologics studied. They serve as examples of good experimental design that will hopefully benefit those interested in planning field trials or in evaluating the results of published trials. General comments about the design of field trials to investigate vaccines against bovine respiratory disease are available also (Martin 1983).

References

- Acres, S. D., and O. M. Radostits. 1976. The efficacy of a modified live Reo-like virus vaccine and an *E. coli* bacterin for prevention of acute undifferentiated neonatal diarrhea of beef calves. *Can. Vet. J.* 17:197-212.
- Armitage, P. 1971. *Statistical Methods in Medical Research*. New York, N. Y.: John Wiley & Sons.
- Burns, K. N. 1963. A comparison of the glucogenic effects of some compounds used in the treatment of ketosis. *Vet. Rec.* 75:763-68.
- Byar, D. P., R. M. Simon, W. T. Friedewald, T. T. Schlesselman, D. L. DeMets, J. H. Ellenberg, M. H. Gail, and J. H. Ware. 1976. Randomized clinical trials: Perspectives on some recent ideas. *New Eng. J. Med.* 295:74-80.
- Cochran, W. G., and G. M. Cox. 1957. *Experimental Design*. New York, N. Y.: John Wiley & Sons.
- Colton, T. 1974. *Statistics in Medicine*. Boston, Mass.: Little, Brown.
- Comstock, G. W. 1978. Uncontrolled ruminations on modern controlled trials. *Am. J. Epidemiol.* 108:81-84.
- Cornfield, J. 1978. Randomization by group: A formal analysis. *Am. J. Epidemiol.* 108:100-2.
- Etherington, W. 1983. The effect of gonadotrophin releasing hormone and/or cloprostenol on reproductive performance in Holstein-Friesian dairy cows: A field trial. M.Sc. thesis, Univ. of Guelph, Ontario.
- Fleiss, J. L. 1973. *Statistical Methods for Rates and Proportions*. Toronto, Canada: John Wiley & Sons.
- Gilbert, J. P. 1974. Randomization of human subjects. *New Eng. J. Med.* 291:1305-6.
- Hill, A. B. 1952. The clinical trial. *New Eng. J. Med.* 247:113-19.
- Martin, S. W. 1978. The design of field trials. *Proc. 2nd Int. Symp. on Neonatal Diarrhea*. Univ. of Saskatchewan, Canada.
- _____. 1983. Vaccination: Is it effective in preventing respiratory disease or influencing weight gains in feedlot calves? *Can. Vet. J.* 24:10-19.
- Martin, W., S. Acres, E. Janzen, P. Willson, and B. Allen. 1984. A field trial of preshipment vaccination of calves. *Can. Vet. J.* 25:145-47.
- Peto, R., M. C. Pike, P. Armitage, N. E. Breslow, D. R. Cox, S. R. Howard, N. Mantel, K. McPherson, J. Peto, and P. G. Smith. 1976. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Brit. J. Cancer.* 34:585-612.
- _____. 1977. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Brit. J. Cancer.* 35:1-39.
- Snedecor, G. W., and W. G. Cochran. 1980. *Statistical Methods*. 7th ed. Ames: Iowa State Univ. Press.
- Thurber, E. T., E. P. Bass, and W. H. Beckenhauer. 1977. Field trial evaluation of a Reo-Coronavirus calf diarrhea vaccine. *Can. J. Comp. Med.* 41:131-36.