Design Considerations in Clinical Trials

Carl S. Ribble

Introduction

New forms of therapy, new therapeutic products, and recently developed vaccines may show outstanding promise based upon laboratory data. A vaccine, for example, may appear to be very immunogenic and effective against a specific disease-causing agent in laboratory tests. However, the same vaccine may not work well in the clinical setting where you intend to use the vaccine. Properly designed and executed clinical trials allow us to evaluate a new therapy or vaccine in the clinical setting, in animals affected by the natural disease. Clinical trials can be an extremely powerful tool for veterinarians; however, emphasis must be placed on the words "properly designed and executed" in the previous sentence, for improperly designed and executed clinical trials are meaningless at best, and destructively misleading at worst.

My purpose in this paper is to explain three key design steps which are essential to any clinical trial: choosing an outcome measure, preventing bias, and establishing the role of chance. To show how a veterinarian would work through each of these steps, a specific example will be used. Imagine you are interested in determining whether a new vaccine effectively reduces bovine respiratory disease (BRD). You intend to test the vaccine in a 6,500 head capacity feedlot you service with a health management program. This particular example is appropriate given the comments of Martin that most BRD vaccine trials reported in the literature have serious design flaws (1). To clarify the approach, assume you intend to have only two groups in the trial, one that is vaccinated with the new BRD vaccine (the test group), and one that is not (the control group). The presence of the control group allows for changes in clinical outcome which are not associated with use of the vaccine (results can always be improved by omitting controls!) (2). Also assume that the analysis will be carried out on individual animals, and herd effects are not an important factor; however, see the paper by Waltner-Toews on why this assumption could be dangerous (3). Realize that the details of the example are presented with the intention of showing how one works through the three key design steps. The authors of *any* published clinical trial should demonstrate they have thought about these three steps. The extent to which they have not will undermine the credibility of their results and conclusions.

Outcome measures

The first step is to state clearly what you want to know — to define the objective of the trial. Asking a general question like, "Does this vaccine work?" is *not* good enough. You must decide precisely how vaccine efficacy will be assessed. One or two measures of outcome must be chosen. The chosen outcomes should be clinically important and meaningful, and they should be subject to objective measurement. Martin has stated that two primary outcome measures should be used with one of these being a production measure of some type (4). At the very least, outcome measures should be *priorized* with a clear primary outcome stated. Do not, for example, monitor six outcome measures and report on those which have statistical significance, because you will have one chance in four of incorrectly concluding the vaccine is effective!

Three key design steps are essential to any clinical trial: choosing an outcome measure, preventing bias, and establishing the role of chance

Consider the outcomes available for assessing BRD efficacy in the feedlot. The outcomes include average daily gain, feed efficiency, serological conversion, morbidity ("first-pulls"), and mortality. The first two measures of outcome are useful when the pen is the unit of analysis, because calves are fed by the pen. Serological conversion is impressive to document but the samples are hard to collect and even harder to interpret — what does "serological conversion" mean clinically? Morbidity is important and fairly easy to measure; however, there can be a significant subjective component to BRD morbidity — the definition of a sick steer can vary noticeably between feedlots, between times of the year, and between treatment crews on the SAME feedlot. Nevertheless, morbidity has far more clinical meaning than serological conversion. Mortality is the most objective measure — it is hard to argue about whether or not an animal is dead! Furthermore, mortality is important clinically and economically. Mortality will be the primary outcome measure and morbidity the secondary outcome measure used to assess BRD vaccine efficacy in our example.

Of the 6,500 calves entering the feedlot then, we plan to vaccinate 3,250 and see if there is a mortality difference between the vaccinates and nonvaccinates. There are three reasons we could see a mortality difference between the groups. First, the vaccine may be truly effective, reducing death loss. Second, a bias may be present in the trial where something other than vaccination status causes a mortality difference. Third, a mortality difference could occur simply by chance. We are only interested in the first cause of a mortality difference. Therefore, we want to reduce the likelihood that either bias or chance will play a role in causing a mortality difference.

Department of Herd Medicine and Theriogenology, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0.

Bias

The bias we hope to avoid occurs where some factor other than the vaccine itself "causes" a difference in mortality between vaccine groups. If there are important biases in the trial, we may conclude the vaccine works, when it *really* does *not* work, or vice versa. There are four key times during trial execution where bias might occur: during 1) the selection of the animals, 2) the follow-up and ancillary treatment, 3) the outcome determination, and 4) final analyses.

We may decide to vaccinate all animals coming from northern auction markets, and not to vaccinate all animals coming from southern auction markets. This certainly would be a simple way of deciding who gets vaccinated. But perhaps calves from southern auction markets suffer more BRD. Even if the vaccine is totally ineffective, more nonvaccinated calves will die because of their source. The difference in source biases the trial, favoring demonstration that the vaccine is effective. To avoid this and other unseen biases, *randomization* of calves to vaccine group is necessary. Randomization tends to remove trial entry biases; furthermore, randomization legitimizes any subsequent statistical testing performed during the analysis stage of the trial (5).

How do you randomize? A random number list from any standard statistics text (6), or generated by a computer software package, can be used to create a randomized vaccination list. The procedure is simple and the number of animals in the two groups will not be radically different (so don't worry about "evening-up" the numbers). In the example, a randomized list would be drawn up before the trial so the processing crew simply has to look at (and check off) the next entry on the list to decide whether a calf gets vaccinated. Effective randomization could even be achieved by flipping a coin and vaccinating all calves that "come up heads".

Bias can also occur during the follow-up treatment and outcome determination stages of a clinical trial. The feedlot treatment crew might treat nonvaccinated calves more aggressively than vaccinated calves, biasing the trial. Similarly, the diagnostic decisions of the veterinarian performing necropsies might be affected by knowing whether a calf was vaccinated or nonvaccinated. Even the statistician or data analyst might be affected by knowing which group received the vaccine. The solution to all these bias problems is "blinding." Attempt to keep the treatment crew, the veterinarian performing necropsies, and even the person analyzing the results, unaware of which group the calves came from.

Chance

How is chance eliminated as a factor in the analysis of the trial results? In reality, chance cannot be totally eliminated as a factor in any clinical trial — you can only state the level at which you accept that chance was an *unlikely* factor.

The vaccine either works, or it does not work. If the vaccine does not work, but I conclude from my trial that it does work, I have made a *type I error* — I have concluded incorrectly that the vaccine is effective (Table 1). The probability of committing a type I error is usually set at 5%, this being "alpha" or the "level of significance." Thus, when performing a (one-tailed) statistical test on a mortality difference seen in the trial, I have a 5% chance of concluding incorrectly that the vaccine is effective. If the vaccine works, but I conclude from my trial that it does not work, I have made a *type II error* — I have concluded incorrectly that the vaccine is ineffective (Table 1). The probability of committing a type II error depends on the size of the trial — the larger the trial, the less likely I will conclude incorrectly that the vaccine is ineffective.

TABLE 1
The two kinds of statistical errors that can occur in a clinical trial

Does the vaccine work "in reality?"	Did the vaccine work in the trial?	Type of error committed
No	Yes	I
Yes	No	II

Unfortunately, type II error is often ignored. This is a critical mistake, for without knowing the probability of committing this type of error, you cannot interpret a negative result. If the clinical trial shows no mortality difference between the vaccinated and control animals, you cannot conclude the vaccine is ineffective. There may have been too few animals in the trial to show that the vaccine was effective — you may not have given the vaccine a fair chance! Increasing the size of the trial might therefore be necessary to improve the power of the trial, the power being the likelihood you will correctly conclude the vaccine is effective.

A *power calculation* should be performed prior to any clinical trial. You can preset the power of the trial and calculate how many animals are needed to achieve that power. Alternatively, you can start with the number of animals you have access to, and calculate the power such a trial would have. Answering the following five questions at the outset of the trial is necessary before you can work through the power calculation (7). This procedure also helps you clarify and finalize the trial design.

1) **What is the *main* purpose of the trial?** To determine if a new BRD vaccine reduces calf mortality in the feedlot.

2) **What is the principal measure of outcome?** Total mortality during the feeding period.

3) **How will the data be analyzed for a treatment difference?** A chi-square test with a 5% level of significance for type I error will be used on the simple comparison of dead vaccinated versus dead nonvaccinated calves.

4) **What type of results does one anticipate in the control (nonvaccinated) group?** Say, based on past experience at the feedlot, we expect 4% mortality in the control group.

5) **How *small* a mortality difference is it *important* to detect and with what *degree* of *certainty*?** Identify the smallest clinical difference of practical value you want to detect. You could argue that any mortality dif-

TABLE 2
Sample size[a] for a mortality decrease from 4% to 3%

		Beta (type II error)		
		0.1	0.2	0.5
Alpha (type I error)	0.10	11,610	8,370	3,645
	0.05	14,175	10,665	5,130

[a]Sample sizes calculated using equation 3.14, where the continuity correction is not employed. In: Fleiss JL. Statistical Methods for Rates and Proportions. New York: John Wiley & Sons, 1981: 41

TABLE 3
Sample sizes[a] for a mortality decrease from 50% to 40%

		Beta (type II error)		
		0.1	0.2	0.5
Alpha (type I error)	0.10	843	608	265
	0.05	1,029	774	373

[a]Sample sizes calculated using equation 3.14, where the continuity correction is not employed. In: Fleiss JL. Statistical Methods for Rates and Proportions. New York: John Wiley & Sons, 1981: 41

ference is relevant and must be detected but this is unrealistic since the trial would have to be infinitely large. Here, let's say we want to see at least a 25% reduction in mortality before it is economically reasonable to vaccinate all incoming calves with the BRD vaccine, and we would like to be 90% sure that this difference was detected as statistically significant.

Power equations can be found in many statistics textbooks which allow you to "plug-in" the information obtained by answering these five questions to determine the number of animals you need (8). Numerous tables have also been produced (using these equations) which can be used to read-off an estimated sample size (9). Time spent consulting a statistician during this design stage of the trial is invaluable!

Working with the five questions as answered for the example clinical trial, Table 2 shows several trial sizes that would be required for a mortality decrease from 4% to 3%. These figures represent the total number of animals that would be required (divided between two treatment groups) for various levels of type I and type II errors. You can see immediately that we are in trouble with access to only 6,500 calves because the specified mortality difference with alpha set at 0.05 and beta at 0.10 would require 14,175 calves. We could still run the trial but if the vaccine really does work (but only reduces mortality to 3%) we will have about a *50% chance* of *incorrectly* concluding that the vaccine is ineffective.

What if a *morbidity* decrease from 50% to 40% had been chosen as the primary outcome measure? Table 3 shows the total number of animals required to demonstrate such a decrease. Note that the decision about the primary outcome measure had a far greater influence on the trial size than any consideration of power or level of significance! To demonstrate a 25% mortality difference requires a trial *14 times larger* than to demonstrate a 20% morbidity difference! The price for a more objective outcome measure is very high in this instance, due primarily to the much lower frequency of mortality compared to morbidity.

By answering the "five power questions" for a variety of potential outcome measures, and weighing the pros and cons of each outcome measure against the trial size necessary for that measure, one attains an understanding of which approach is best. An informed decision can then be made about whether to continue with the trial, to modify the trial, or to forget the trial altogether.

References

1. Martin SW. Vaccination: is it effective in preventing respiratory disease or influencing weight gains in feedlot calves? Can Vet J 1983; 24: 10–19.
2. Bearman JE, Loewenson RB, Gullen WH. Muench's postulates, laws, and corollaries or biometricians' views on clinical studies. Bethesda, Md: Biometrics note 4, National Eye Institute, 1974. Quoted in: Bulpitt CJ. Randomized Controlled Clinical Trials. Boston: Martinus Nijhoff Publishers, 1983: 63.
3. Waltner-Toews D. Herd effects in trials of animal biologics. Can Vet J 1989; 30: 000–000.
4. Martin SW. The design of field trials. Proc 2nd International Symp Neonatal Diarrhea, Saskatoon, 1978: 489–507.
5. Bulpitt CJ. Randomized Controlled Clinical Trials. Boston: Martinus Nijhoff Publishers, 1983: 44.
6. Snedecor GW, Cochran WG. Statistical Methods, 7th ed. Ames, Iowa: Iowa State University Press, 1980: 463–466.
7. Pocock SJ. Clinical Trials: A Practical Approach. New York: John Wiley and Sons, 1983: 123–124.
8. Fleiss JL. Statistical Methods for Rates and Proportions. New York: John Wiley & Sons, 1981: 38–42.
9. Fleiss JL. Statistical Methods for Rates and Proportions. New York: John Wiley & Sons, 1981: 260–280.