How to Conduct and Interpret Field Trials

Hollis N. Erb, *D.V.M., Ph.D.*
Department of Clinical Sciences
New York State College of Veterinary Medicine
Cornell University
Ithaca, NY 14853

Some bovine practitioners are beginning to conduct their own field trials in order to test new drugs, surgical techniques, feed additives, management programs, and the like. Other practitioners are being asked by scientists in universities or industry to participate in trials, and all practitioners should be reading reports of field trials in journals. There is a science to field trials. Trials that are bad science shouldn't be conducted or believed.

In this essay I will discuss 6 of the major issues in the design of field trials: controls, sample size, informed consent, baseline measurements, assignment of treatments, and blindness. I will not discuss statistical issues other than some general comments regarding sample size. I'll generally take the point of view of the practitioner trying to design his or her own trial, but these same issues may be used to evaluate the quality of other people's trials. (For further reading, there's an excellent recent article by Dohoo and Waltner-Toews (1), and a very good, non-statistical introductory textbook by Friedman et al. (2).)

Controls

All biologic systems are extremely variable, and you almost never can predict with certainty "what would have happened" had you not intervened with the new, experimental treatment. Therefore, there must be a control group. The rare exceptions are the instances in which there was absolutely no doubt about the outcome and time course. No convenient veterinary examples come readily to mind; the classic human example in which no controls were needed in order to demonstrate effectiveness was the introduction of streptomycin for the treatment of tuberculous meningitis.

Concurrent Controls Are Best

The best controls are involved in the trial at the same time as the subjects who are receiving the new treatment. The reason that "historic controls" are rarely as satisfactory as "concurrent controls" is that so many factors vary across time. Winters vary in severity, insect vectors wax and wane, steers are purchased from different sources, etc. Very important is the fact that quality of records varies and memories of what you *think* happened last year is extremely fallible. (It is a rueful joke among researchers that a disease always "dries up" when you set out to study it formally. This happened to me when we estimated (with help from the clinicians servicing the farms) that local dairy heifer calves would experience a 30% incidence rate of calf scours; when we established criteria and started keeping records, the rate was closer to 15%.) The point to take away is that the best field trials use concurrent controls.

What Treatment is Given to the Controls?

Most people have a sort of knee-jerk reaction and would answer that, "Controls get a placebo, of course." In fact, what's done to the controls depends on the research situation and the research objectives. The options are: "standard therapy," "placebo," and "nothing."

If you are testing an experimental *therapeutic* regimen, then the proper control treatment is standard therapy (assuming one exists). To leave your client's livestock untreated when a standard therapy for their illness exists would be unethical. Besides, if the new therapy is to be worthwhile, it must prove to be more useful than the treatment already in common usage—not better than saline.

If there is not already a standard therapy, then the objectives of the field trial dictate whether the controls get a placebo or are left alone. The criterion is whether the intention is to test the "active ingredient" or whether you mean to test the whole regimen—active ingredient *plus* the process of delivering it (a "management" or "cost-benefit" trial). The active ingredient trial might ask whether the new PI3 vaccine will raise serum antibody levels. The management type might ask whether the vaccine protects well enough to be worth the costs of vaccine, labor of the cowboys, and pushing the steers away from the feed bunks and through the chutes (and running the risk of a steer breaking his leg). An active ingredient trial requires a placebo for the controls, so that the hassle of delivering the experimental treatment is similar for both groups. On the other hand, the management or cost-benefit type trial calls for a "do nothing" control treatment. (Frankly, the management type trial is probably of greatest interest to clients and is the real measure of potential acceptance of the new treatment.)

In summary, any field trial that does not include controls probably is bogus. The best controls are concurrent rather than historic. There should not be a uniform insistence on

placebo controls; rather, the new treatment should be tested against the treatment it is intended to replace. In some instances, this will mean that controls receive standard therapy; in other instances, it is appropriate to leave the controls alone.

Sample Size

Sample size needs should be considered early in the planning of a trial. (Formulas or tables are available in most statistics texts.) Larger sample sizes will be needed than you probably imagine. For instance, approximately 500 calves per group (1,000 calves in all) would be needed in order to tell that an incidence rate of pneumonia in veal calves of 5% was significantly different from 10%. A similar sample size would be needed for dairy cows with 8,300 kg records, if you wanted to tell whether the new feed additive could increase yield by about 250 kg (3%).

In general, larger sample sizes would be needed if:

- the outcome is rare rather than common;
- the variability in the outcome is large rather than small;
- it would be important to detect small rather than only large differences;
- losses-to-followup are anticipated; and
- mistaken conclusions from the trial are relatively costly.

("Mistaken conclusions" means saying that the treatments are different when they really aren't, or saying that the treatments aren't different when really they are.) Unfortunately, many published reports are of trials that were too small to detect differences that would in fact have had clinical or practical importance. This is a waste of resources and a disservice to clients and to the profession. In fact, it could be argued that it is unethical to put a client's animal *at risk* in a trial that is too small to be useful. (When reading published reports, don't criticize small trials that *did* find significant differences, but if the trial found "no significant difference," consider whether there could be a sample size problem.)

There is an additional warning regarding sample size. Sample size is the count of the units to which the treatments are assigned. If you are testing a new management policy that must be applied on a "whole-herd" basis, then "herd" is 1 sample size unit—*not* the number of cattle in the herd. This is because the individual cattle within the herd are more alike than a mix of cattle from different herds, so that if all the cattle within the herd *have* to get the same treatment, they don't count as individual, independent experimental units. For instance, if you test different mechanical ventilation rates in different rooms of a veal calf building, the sample size technically is based on the number of rooms rather than on the number of calves in each room. (Unfortunately, this has been a common source of error in the past.)

Informed Consent

At this point, you should have an idea of your field trial's design from the point of view of the treatments that will be tested, the type of subjects you need, and how many subjects you need. Now you have to recruit those subjects, which means you have to go and ask the herd managers for permission to run the trial in their herds.

This process of asking permission means getting "informed consent." At the bare minimum, what you have to tell the herd manager includes:

- the objectives of the trial;
- the methods that will be used, including the alternate treatments;
- the risks (those that reasonably can be anticipated);
- the benefits (those that reasonably can be anticipated);
- the fact that the herd manager may decline to participate (and that declining won't in any way damage your client-veterinarian relationship); and
- the fact that the herd manager may withdraw from the trial at any time (again, without harm to your professional relationship).

Informed consent is not at this time a legal requirement. However, I recommend to you most strongly that you get informed consent before you proceed, for 2 reasons. The 1st reason is that the client has an economic investment in the experimental subjects and therefore has a *right* to decide whether they will be used in this fashion. The 2nd reason is a form of self-defense. Field trials are done to uncover *risks* associated with new treatments—not just to document benefits. If something does go wrong during the trial, I believe you will be in a much happier position if the client was fully informed in advance about the research, rather than having the client discover only after the fact that you were "experimenting" with his animals.

(Informed consent may sound complicated, but in my experience with 2 field trials and several years of observational studies on dairy farms, farmers are *extremely* cooperative and almost never decline to participate. Informed consent doesn't "scare them off.")

After the informed consent process, you will have available a pool of subjects who have been volunteered (by the herd manager) for the trial. There will be a temptation which you *must* resist to give the experimental therapy of the volunteers, and to use the *non*-volunteers as controls (in a "standard therapy" or "do nothing" comparison). The potential for bias if you do this is so great that it will invalidate the study. The control and experimental subjects must be alike in all ways *except* for the intervention being tested, and volunteers *a priori* always are presumed to differ from non-volunteers. Careful scientists will reject your research if you make this mistake.

Baseline Measurements

Before you begin treatments in your trial, you must take pre-treatment, "baseline" measurements. Not as obvious is that in many instances these measurements should be taken before you decide which animal is to get which treatment.

When to take Baseline Measurements

There are 2 reasons to take baseline measurements *before*

assigning treatments. The first reason is that if you don't yet know the treatment assignments (if you are "blind"), then that knowledge can't bias (influence) the readings on the measurements. This is especially important if the "measurements" are at all subjective. ("Blindness" is discussed in the last section of this paper.)

The second reason to take baseline measurements before assigning treatments is to ensure that the subjects do in fact qualify for admission to your trial. It may make no sense, for instance, to test a new IBR vaccine in steers that already have titers.

What Baseline Measurements to Take

There are no hard-and-fast rules for identifying which variables should be measured at baseline. However, a good rule-of-thumb is to take measurements on the *outcome* variable (the thing that might change as a result of the experimental treatment) and on important *confounding* variables. First, these variables probably will be important (as mentioned above) in determining eligibility of the animal or herd for the trial. Second, you may wish to make use of these measurements to form groups of similar subjects, and then assign treatments evenly within the similar groups. For instance, pairs of cows with negative milk cultures (entry criterion) might be matched on parity and somatic cell count (SCC) before beginning a trial on mastitis prevention. Then, if 1 member of each matched pair is used as a control and the other cow is given the experimental program, you will know that there can be no differences between the treatment groups regarding parity, nor can SCC confound (bias) the results of the trial.

The 3rd reason to have pre-treatment measurements on the outcome and confounding variables is that your assignment procedure may *not* have created equivalent groups. With perhaps a little help from a statistician, adjustments for the differences can be done during the statistical analysis—but you have to have the baselines to do so. Finally, the 4th reason is that the baseline measurements will (with luck) show that your treatment groups really were roughly equivalent at the start. You can't convince other people of this unless you have the measurements to show them. (Also, of course, any evident differences will explain to other scientists the reason for the adjustments to counteract confounding that you might have made in the statistical analysis.) For these latter reasons, high-quality articles reporting the results of field trials will include summaries for each treatment group of the baseline measurements.

Treatment Assignment

Randomization

A biased treatment assignment procedure is a "fatal flaw" in the design of a field trial. The mere possibility that the procedure *might* have been biased will irrevocably damage the credibility of your work.

The best way to assure that the assignment procedure wasn't biased is to do a *randomized* trial. Randomization means that the treatment assignments were based on a *definable,* fair mechanism that relies on Chance (comparable to the classical "drawing lots" or "tossing a coin"). Looking at each heifer in turn and deciding "on impulse" whether or not she will be a control is not a random assignment—it is a "haphazard" assignment. Haphazard assignment is terribly subject to bias, and completely invalidates any trial. Because this issue is so important—and because many lay people mistakenly use "random" as a synonym for "haphazard"—many authors will make a point of indicating the chance mechanism that was used (e.g., a "table of random numbers").

Reasons to Randomize

One of the reasons for randomizing already has been mentioned: randomization guarantees against the scientist imposing a bias during the assignment process. The second reason also is for the ultimate purpose of controlling bias—the bias introduced by confounding variables.

As mentioned above, it is possible to use baseline measurements to force the groups to be balanced on important confounding variables. However, this method of bias control has 2 limitations. You can match on only a small number of variables (perhaps 2 to 4), because it quickly becomes impossible to find simultaneous perfect matches on *each* of several factors. More importantly, you cannot match on confounding variables that can't be measured conveniently at baseline, and you can't match on confounders that are not even suspected. The *only* way to guarantee that *most* of the unknown, unsuspected confounding variables will be balanced *approximately* equally between treatment groups is to randomize the assignment. ("Most" and "approximately" are important; randomization doesn't guarantee *perfect* equality, which is why some adjustments for baseline measurements of known confounders still may have to be done in the statistical analysis in spite of the use of randomization.)

Systematic Assignment

There is one additional way to make treatment group and assignments: systematic assignment. Systematic assignment is assignment based on a *definable,* regular *alternating* pattern. Examples might be "every other calf through the chute" or "cows with odd (versus even) DHI cow control numbers."

If no confounding variable is distributed in a similar pattern and *if* such a procedure doesn't unblind a trial that needs to be blind, then systematic assignment might be "about as good as" randomization. *The burden of proof is on the researcher,* though, to present a convincing argument.

The systematic assignment must be based on a repeating, alternating pattern. If 100 calves are boiling out the pen through the chute, alternating calves are equally likely to be the biggest, the weakest, the most sick, the most hungry, the

most frightened, etc. However, it would be easy to suspect that the 1st 50 out of the pen might *as a group* be different from the last 50. Taking the 1st 50 as one treatment group and the last 50 as the other would be biased and would *not* qualify as an alternating, acceptable systematic assignment.

The best way to assign treatments is to do so randomly. If, instead, systematic assignment is used, it must be *called* "systematic" in the resulting publication. The researcher should explain in the article why it was an acceptable method under the circumstances of the particular trial, and also must be especially careful to present the data regarding pre-treatment measurements.

Blindness

In Measuring the Outcome

Under the best circumstances, the person measuring the outcome at the end of the trial will be "blind" . . . that is, ignorant of the treatment group of the animal being examined. Such blindness is an assurance that the measurement will not be biased in favor of either treatment.

Field trials should be blind if at all possible, but sometimes it *isn't* possible for practical or ethical reasons. For instance, you can't hide the fact of a bandaged foot in a trial that's testing bandaging versus leaving the foot unbandaged, and you can't impose sham surgery on a client's cattle. Under these (unblind) circumstances, great care must be taken to assure that the outcome measures are objective and free of any reliance upon subjective interpretations. The foot bandaging trial actually was done (3) and correctly limited the diagnosis to 3 defined categories (essentially, "not lame," "lame but weight bearing," and "not weight bearing"). Any more complicated scoring would have been challengeable on the basis of subjectivity (therefore, possible bias in favor of either treatment group) and also, incidentally, lack of repeatability between different observers.

Blindness During the Trial

It is conceivable that herdsmen might give preferential treatment to animals in one or the other group, if the herdsmen knew which animals were in which group (and which group got which treatment). Therefore, the herdsmen also should be kept blind during the trial, if it's possible. Blindness certainly *is* possible if the controls are given a a placebo (and sometimes even if they are given standard therapy). Under placebo-controlled trials, blindness of the herdsmen should be maintained.

When publishing the results of a trial, the methods used to guarantee blindness or the justification for non-blindness (and the measures used to protect against the bias due to the lack of blinding) should be described. It is especially important to discuss the blindness of the person measuring the outcome (because non-blind subjective assessments are a fatal flaw), but blindness of the herdsmen also should be addressed.

Conclusion

Most of the major non-statistical issues regarding field trial design and evaluation have been discussed. A statistician should be consulted for help in analyzing the data. The statistician also can be helpful in determining the appropriate sample size, the form in which measurements should be recorded, and the actual process of randomizing the study subjects. For these reasons, the statistician should be consulted *before* you begin your trial. (I suggest you try to find a consulting statistician through your land grant university.)

There is no excuse for a field trial with inappropriate controls, biased treatment assignment, or non-blind subjective evaluation of the outcome. These errors are fatal flaws, and any one of them is sufficient to invalidate the trial. When you plan, discuss, or read of the design of a trial, check these issues first. If you are satisfied with the controls, randomization, and blindness, then go on to examine the issues of sample size and baseline measurements. Small sample size will warn you that small but important effects of treatment might be missed, and the baseline measurements will warn of potential confounding to consider when interpreting the results.

Failure to acquire or to mention informed consent does not invalidate the scientific quality of the data, but in the near future it will be a hallmark of an ethical researcher.

References

1. Dohoo, I.R. and D. Waltner-Toews. Interpreting clinical research. Part I. General considerations. Compend. Cont. Ed. 7:S473-S478. 1985. 2. Friedman, L.M., C.D. Furberg, and D.L. DeMets. *Fundamentals of Clinical Trials.* Second edition. PSG Publishing Company, Inc. Littleton, Massachusetts. 1985. 3. White, M.E., L.T. Glickman, I.C. Embree, P.M. Powers, and E.G. Pearson. A randomized trial for evaluation of bandaging sole abscesses in cattle. J. Am. Vet. Med. Assoc. 178:375-377. 1981.