Sponsors

University of Minnesota
College of Veterinary Medicine

College of Agricultural, Food and Environmental Sciences

Extension Service

Swine Center

Editors
W. Christopher Scruton
Stephen Claas

Layout
David Brown

Logo Design
Ruth Cronje, and Jan Swanson;
based on the original design by Dr. Robert Dunlop

Cover Design
Sarah Summerbell

Evaluation of clinical trials: A practical approach to distinguishing the foo-foo dust from reality

Gail Cunningham BS, DVM, MVetSc
Marshall Swine Health Services, 4712-41st Camrose, Alberta, Canada, T4V 0Z6, Gailc@cable-lynx.net

Introduction

In veterinary medicine today, we are continually bombarded with information on new products, procedures, and treatments. It is a challenge to keep up with all of the new information in any area of veterinary medicine, much of which comes in the form of randomized controlled clinical trials.

A randomized controlled clinical trial is defined in the following way:

> An epidemiologic experiment in which subjects in a population are randomly allocated into groups, usually called study and control groups, to receive or not to receive and experimental preventive or therapeutic procedure, maneuver or intervention. The results are assessed by rigorous comparison of rates of disease, death, recovery or other appropriate outcome in the study and control groups respectively. Randomized controlled trials are generally regarded as the most scientifically rigorous method of hypothesis testing available in epidemiology. It may also be referred to as a "Randomized control trial" (1).

Randomized controlled clinical trials are used to evaluate medications, vaccines, surgical procedures, feed additives, management tools, and many other areas of practice. Properly designed and executed trials allow us to evaluate these new products and processes in a clinical setting in animals that are, or should be, affected by natural disease. The key here is "properly designed and executed." Part of the job of an epidemiologist is to evaluate the design and execution of these trials to ensure that the results obtained are "true" or "real." We need to be able to evaluate the role of bias and chance in the outcome of the trial. I will offer one more definition before I continue:

"Bias: Any trend in the collection, analysis, interpretation, publication or review of data that can lead to conclusions that are systematically different from the truth" (1).

Bias and chance can be minimized in any trial by good design and analysis. Randomization of subjects; blinding of subjects and, possibly, the observer; adequate sample size; control of potential confounders; and statistical analysis are all key elements of a good clinical trial. As practitioners, we need a simple but methodical way to evaluate these elements of a trial. That's where those epidemiologists fit in!

As a graduate student in epidemiology at the Western College of Veterinary Medicine, I was taught ten questions to use in the critical evaluation of a clinical trial paper. I have used this method almost daily in practice and in the evaluation of the design of my own research projects. They provide a practical framework for quick but effective evaluation of the clinical trials that we use for continuing education in our practices.

Criteria for the critical evaluation of a clinical trial paper

These criteria were developed by Drs. Cheryl Waldner and John Campbell of the Department of Large Animal Clinical Sciences at the Western College of Veterinary Medicine, University of Saskatchewan.

Was the assignment of animals to treatment groups randomized?

Examine the title, abstract, materials and methods, and statement of objectives. The study should mention some kind of randomization procedure (randomized, random allocation, randomly assigned) not simply state choices were made "at random." Randomization minimizes selection or entry bias (for example, variation in disease severity between treatment and control group). If the assignment is not truly random, there is much increased risk for bias and the role of chance in any reported outcome.

Ask yourself, were the control groups concurrent or historical? Concurrent control groups are formed at the same time as the treatment group. If historical or previous control groups are used, too many factors unrelated to treatment can produce any differences that may be observed. There are differences in housing, environment, weather, feed, health status, etc. from past to present. The use of historical controls tends to exaggerate the value of a new treatment because of variation in health and productivity from year to year.

How were vaccine and treatment trials handled?

Specifically—how were the trial animals challenged and did disease occur in the controls? Evaluate whether it was a laboratory challenge or a natural challenge for both the treatment and control animals. Also evaluate whether or not any of the control animals became sick. Some of the controls should get the disease in question. If they do not, there was likely not sufficient challenge to evaluate the efficacy of the vaccine.

Also ask yourself whether the case definition (inclusion criteria) was clearly stated and defined. There should be a precise case definition for the condition on which the treatment is being assessed, as well as documented diagnostic criteria for inclusion and exclusion.

Were all clinically relevant outcomes reported?

Assess whether all animals who entered the study were accounted for at its conclusion. Count the number of experimental and control subjects at the beginning and the end of the trial. These numbers should be *identical*. Death from all causes should be reported—not just those from the disease of interest. Examine the results to see whether mortality as well as morbidity was reported. Finally, ask yourself whether the issue of quality of life (such as in cases of cancer treatments for pets) was addressed.

Was the measure of outcome meaningful?

Was the outcome clinically important? Was serology used to evaluate disease, morbidity, or mortality? The outcome should be reasonably measurable and provide clear evidence that the treatment works.

Were the biology and the epidemiology of the disease considered?

For example, is there a specific age or sex of animal that the disease affects? You would not evaluate a vaccine for neonatal diarrhea on the finish floor of a swine production unit. The incubation period of a disease and the time required for an immune response from a vaccine should be considered. If the animals are vaccinated and their disease starts three days later, the vaccine has not had time to generate an immune response.

Were blinding techniques used to reduce bias?

These techniques include the following:

- Single blind: The administrator of the treatment (patient, owner, client, barn worker)
- Double blind: The administrator of treatment and measurer of the outcome (the investigator and the patient, owner, client, worker)
- Triple blind: The administrator of treatment, the measurer of outcome, and the analyzer of results (that is, the person doing statistical analysis is also blind)

Drugs under investigation should look identical or be masked. Blinding is not always possible, such as when you are comparing two very different treatments such as chemotherapy and surgery. Blind analysis of the results after treatment, however, can still be done.

What other potentially important biases are evident?

Pay attention to the trial specifics and design characteristics. Did they identify specific potential biases, and are they enough to discredit the entire paper? All field trials will contain some bias and the authors should identify and discuss some potential biases.

Were contamination and co-intervention avoided? Contamination occurs when controls accidentally receive (some element of) the treatment. This might decrease the difference seen between treatments and controls. Co-intervention occurs when there is performance of additional diagnostic or therapeutic acts on the experimental but not control subjects. This will increase the difference seen between the treatment and control groups. This can be prevented by blinding.

Was compliance measured? This can be important if owners are giving animals daily medication or other treatments that must be given and monitored. It is important that this compliance be measured in the analysis by such means as tablet counts, serum levels, or other relevant biomarkers.

Were withdrawals, non-compliers, and those who crossed-over handled appropriately in the analysis? Withdrawals can be losses due to follow-up. These can include animals that died because owners or participants stopped giving the medication or treatments. It may also include animals that became ill and required additional treatment. These should be included and handled appropriately in the analysis; they cannot just be skipped or left out.

How likely was the result a chance finding?

Statistical significance tells us whether the conclusions are likely to be true, regardless of clinical significance (the P-value of 0.05 for statistical significance). If the observed difference was statistically significant, was it clinically important? Some results that are not statistically significant can still be clinically important. The reverse is also true. Statistical significance does not always mean that the result is clinically relevant (e.g., reduction in relative risk preventable fraction).

If results were not statistically significant, assess whether the study was big enough to show a clinically important difference, if it should occur. If power is very low, then the type II error is very large. This means that with a small sample size and low power, we can easily miss a treatment difference when one actually exists. The number of subjects should be determined at the same time as the

primary research end-point. Power is usually set at 80%, meaning that 20% of the time we will miss a difference that exists. Ninety percent power is also common in the literature.

What are the differences between the trial animals and animals in your practice?

Are the animals in the study so different that you cannot apply the study results in your own practice? This includes differences in environment, housing, feeding practices, genetics, etc. The animals must show enough similarity for the results to have meaning in your situation.

Also consider whether reproducibly defined exclusion criteria are stated. Inclusion criteria define why some animals are left out of the trial. If too many exclusion criteria are used, then the generalizability of a trial is compromised. Too many exclusion criteria means it is targeted at a very specific population only.

Is the therapeutic maneuver feasible in your practice?

Was the therapeutic maneuver described in sufficient detail for you to reproduce it? Is the formula, dose, method of administration available, affordable, and sensible? Can it be put into use in a practical manner in your current practice or situation? There are also specific considerations in trial design that should be addressed in the final presentation, such as specific web sites used for sample size calculations and epidemiological software tools that are documented and available for download off the web.

Conclusion

Correct evaluations of the clinical research we read daily is extremely important to sort out the research that is meaningful from that which is not. By following the above steps, you can quickly evaluate if a specific trial is useful and meaningful to you. This heuristic provides us with a more efficient and methodical approach to separating the "foo-foo dust from reality."

References

1) Last JM ed. *A Dictionary of Epidemiology,* 3rd ed. New York: Oxford University Press, 1995.
2) Hennekens CH, Buring JE. *Epidemiology in Medicine,* 1st ed. Boston: Little, Brown and Company, 1987.
3) Fletcher RH, Fletcher SW and Wagner EH. *Clinical Epidemiology: The Essentials* 2nd ed. Baltimore: Williiams and Wilkins, 1996.
4) Ribble CS. Design considerations in clinical trials. *Can Vet J.*