Clinical trials: How many animals do you need ?

Dirk Pfeiffer

Department of Veterinary Clinical Sciences, Massey University, Palmerston North, New Zealand

Clinical trials

The objective of a clinical or intervention field trial is to provide information on whether a particular treatment or change in management practice (treatment) has an effect with regard to a particular disease or production problem (outcome variable). Hence, there will be a group representing the current situation i.e. no intervention (controls) and one or more groups (actual treatments) where a treatment has been applied or a particular management practice has been changed. At the end of the trial there may or may not be a difference in disease incidence, mortality or a production parameter between the control and treatment groups. The investigator then has to make a decision on whether the treatment has an effect or not. In statistical terminology the term *null hypothesis* says that there is no difference between the groups and the *alternative hypothesis* suggests that there is a difference between groups. This decision would be very easy if there was no variation between individual animals in the response to the treatment, in the untreated control animals and in the particular outcome variable. Unfortunately this is not how nature works, instead we have to deal with the probability of certain outcomes and responses. Statistical methodology is the appropriate tool for dealing with these situations. If there was a difference, statistical analysis techniques can be used to assess how likely it was that the difference was due to random variation. If there was no difference, the question has to be asked how likely it was that a difference would have been detected if it had been there. Both questions relate to the size and variation of the actual difference found between treatment and control groups, but also to the sample size which had been used.

Estimation of sample size

In order to decide on the required sample size during the design phase of the study the researcher has to establish what *magnitude* of difference between treatment and control groups in the outcome variable would be meaningful based on economic, animal or public health considerations. The difference could be measured as the difference between proportions (such as mortality) or between means (such as average milk production). For a continuous type variable the *standard deviation* of the outcome variable in the population investigated is also required. This can be derived from previous work, the literature, a pilot study or from an educated guess. It then has to be decided what chance of detecting such the expected difference is required. This is called the statistical *power* of the study. Typically, an 80% chance of detecting a difference, if it is there, is desirable (*note*: *beta* or *type II error* equals (1 - power)). More recently it has beeen advocated that the *beta error* should be the same as the

alpha error. The next decision will be on what probability the researcher is prepared to accept that the observed difference was not a true difference, but due to random variation alone. This is also called the *alpha* or *type I error*. Typically an *alpha error* of 0.05 or 0.01 is selected. As soon as these decisions have been made the required *sample size* can be estimated.

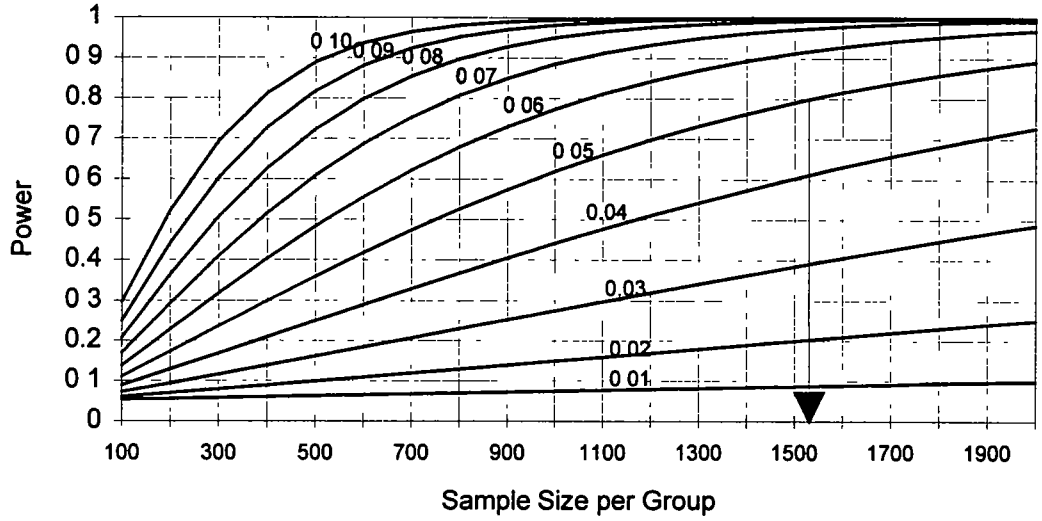
The following relationship between the four parameters *sample size*, *magnitude of the difference*, *alpha* and *beta error* exists:

The required *sample size* increases given
a decrease in the *magnitude* of the difference or
an increase in the *variability* of the outcome variable or
an decrease in the probability of not detecting a difference which is actually there (*beta error)* or
a decrease in the probability that the difference was due to chance variation (*alpha error*).

Example for a categorical type outcome variable (conception yes/no):

The objective is to test the effect of gonadotrophin-releasing hormone treatment at the time of insemination on conception (source of data: Morgan,W.F. and Lean,I.J. 1993: Gonadotrophin-releasing hormone treatment in cattle: a meta-analysis of the effects on conception at the time of insemination. *Aust.Vet J.* 70(6), 205-209). Available data from the literature suggests that conception probability of cows at the time of first insemination after calving is about 0.40. One group of cows will be treated and a second group will be used as controls. Given the cost of the treatment a difference of 0.05 between conception probabilities of the treatment and control group is considered economically desirable. A statistical power of 80% and a beta level of 0.05 will be the basis for estimating the required sample size. Figure 1 shows a graph of the different sample sizes required to achieve a certain power with the lines representing different magnitudes of difference between treatment and control groups. A sample size of 1525 animals in each group would be required to have an 80% probability of detecting a difference in conception rate of 0.05 using a p-value of 0.05.

Figure 2: Power and sample size for different expected differences in conception probability between the two groups given an alpha error of 5%



Example for a continuous type outcome variable (serum copper level):

The effect copper supplementation on growth rate in severely copper deficient weaner deer is to be evaluated using a clinical trial. The outcome variable is weight gain mesured over a period of 12 months. The trial population consists of weaner deer of both sexes with serum copper levels of less than 8µml/l. Over the 12 month period a difference in average weight gain of 3kg between the two groups is expected. The standard deviation for weight gain in the control group is estimated to be around 3kg. A *power* of 80% and an *alpha error* of 5% is used to estimate the appropriate *sample size*. Figure 3 shows the relationship between *power* and *sample size* for different *magnitudes* of the difference in average weight gain between the two groups. A *sample size* of about 16 animals per group would provide 80% *power* at an *alpha error* of 5% and assuming a *standard deviation* of 3kg in weight gain. If a *power* of 95% were required, the *sample size* would have to be about 27 animals per group. A reduction in the expected difference between average weight gains to 2kg would result in an increase of *group size* to about 35 animals. Figure 4 demonstrates the effect of varying the *standard deviation* of weight gain in the groups on power and sample size at an *alpha error* of 5% and a expected difference between average weight gains of 3kg. If the *standard deviation* increases to 5kg, to achieve 80% *power* group size has to be about 43 animals. If the *standard deviation* were only 2kg, a group size of about 7 animals would provide 80% *power* at an *alpha error* of 5% and an expected difference between average weight gains of 3kg.

Figure 3: Power and sample size for different expected differences in average weight gain (1 kg to 5 kg) based on an alpha error of 5% and a standard deviation of weight gain of 3kg

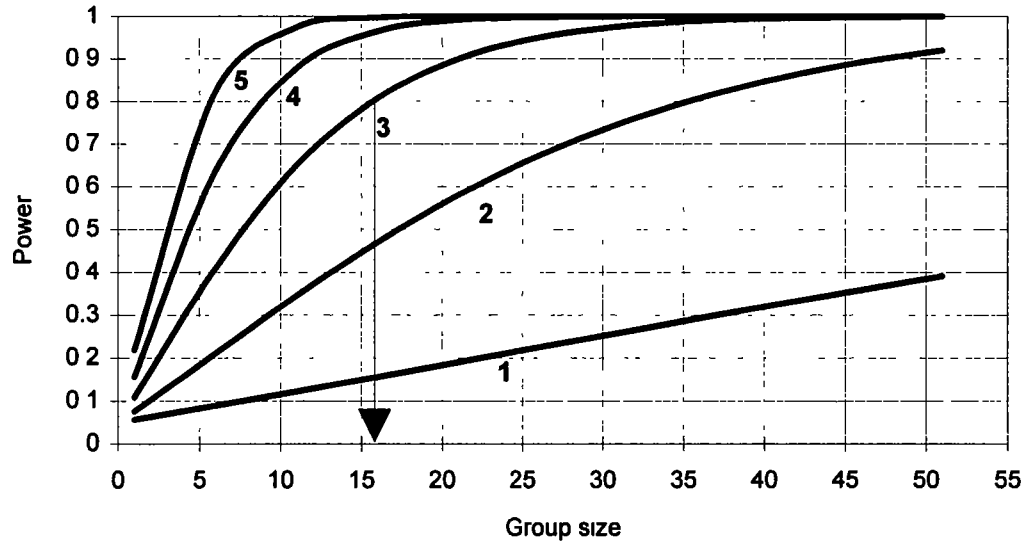
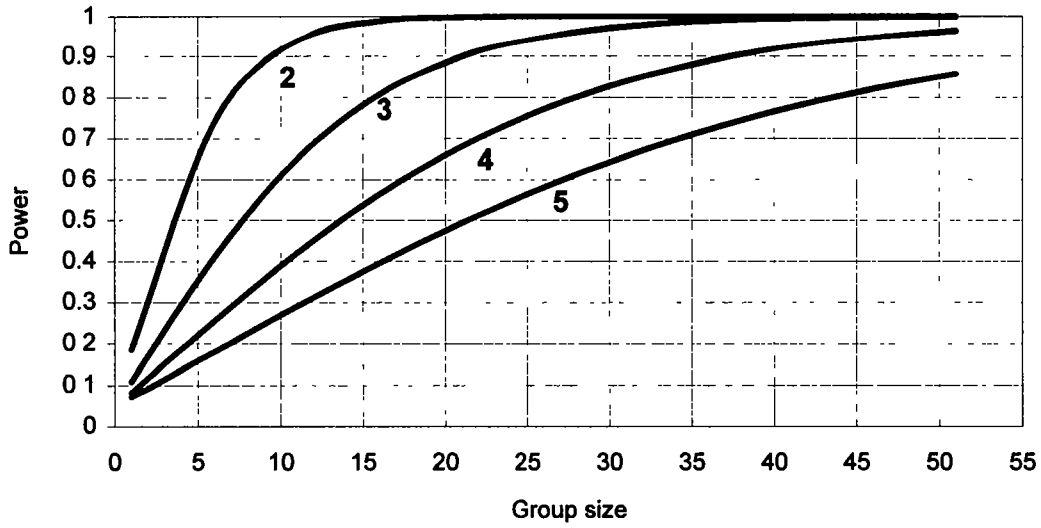


Figure 4: Power and sample size for different standard deviations of weight gain (2 to 5) at an expected average difference in weight gain of 3kg and a alpha error of 5%



Statistical significance and the outcome of a clinical trial

The analysis of the data which was collected during a clinical trial will provide an estimate of statistical significance for the observed difference between the two treatment groups. But this does not necessarily provide the answer to the question which was asked at the beginning of the trial: Does the treatment have an effect or not ???

Assuming that the result of the analysis attributes a statistical significance level of *say* $p = 0.01$ to the observed difference between groups, the following conclusions can be made:

- If there was no true difference between the two groups, then during a series of *say* 100 trials just one time (0.01 * 100) a difference of the magnitude of the observed difference would have been found just by chance.
- In statistical terminology this means that the *null hypothesis* can be rejected in favor of the *alternative hypothesis* at the significance level of 0.01.
- **But** it cannot be concluded that there is a 99% chance that the observed difference between groups truely exists.
- Statistical significance is not synonymous with biological, clinical or economical relevance.
- **Remember**, that with a large enough sample size it may be possible to show that even small practically unimportant differences are statistically significant.

If the result of the analysis was not statistically significant at a given level, the following conclusions can be made:

- There is a chance greater than the specified significance level that there is no difference in the outcome variable between the two groups.
- In statistical terminology the statistical test fails to reject the *null hypothesis*.
- Depending on the sample size the power of the analysis could have been too low to detect an existing difference.

The decision on whether the treatment had an effect or not has to be made on the basis of these statistical results **and** other information such as problems occurring during conduct of the trial, costs, risks, consequences of the outcome etc..

Where to get this information from ?

As long as the researcher has decided on the magnitude of the difference, the statistical *power* and the *alpha error* required, most statistical books include the tables which will provide information on the appropriate *sample size*. There are also a number of computer software packages available which can be used to calculate the required *sample sizes*. The following are some examples:

- STATCALC (USD, Inc, 2075A West Park Place, Stone Mountain, GA 30087) - a standalone which is part of the public domain epidemiological analysis software Epi Info - can be used for power analysis of dichotomous type (yes/no) outcome variables as well as for performing chi-square analysis of tables
- PASS version 1.0 (NCSS, 329 North 1000 East, Kaysville, Utah 84037, Fax: 801 - 546 3907) allows the estimation of power and sample size for a number of different statistical analysis problems.
- STPLAN 4.0 (Barry W. Brown, Dept. of Biomathematics, Box 237, University of Texas, Houston, TX 77030, U.S.A.) is in the public domain and supports an even

more complete set of different study designs, but is more difficult to use and does not provide graphical output of the effects of varying the different parameters.

- PC-Size (Gerard E. Dallal, 54 High Plain Road, Andover, MA 01810, U.S.A.) is a shareware program which allows estimation of power and sample size.