Sample Size Determination

Ralph B. Dell, Steve Holleran, and Rajasekhar Ramakrishnan

Abstract

Scientists who use animals in research must justify the number of animals to be used, and committees that review proposals to use animals in research must review this justification to ensure the appropriateness of the number of animals to be used. This article discusses when the number of animals to be used can best be estimated from previous experience and when a simple power and sample size calculation should be performed. Even complicated experimental designs requiring sophisticated statistical models for analysis can usually be simplified to a single key or critical question so that simple formulae can be used to estimate the required sample size. Approaches to sample size estimation for various types of hypotheses are described, and equations are provided in the Appendix. Several web sites are cited for more information and for performing actual calculations.

Key Words: number of animals; power calculation; sample size; statistical analysis

In the United States and in most European countries, an investigator must provide the animal care committee with an explanation for the number of animals requested in a proposed project to ensure appropriateness of the numbers of animals to be used. This article is written for animal care committee members and veterinarians and for researchers who are asked to provide statistical calculations for the proposed number of animals to be used in their project. The project's purpose may be to obtain enough tissue to do subsequent analyses, to use a small number of animals for a pilot experiment, or to test a hypothesis. In the text below, we discuss the statistical bases for estimating the number of animals (sample size) needed for several classes of hypotheses. The types of experiments that an investigator might propose and the methods of computing sample size are discussed for situations where it is possible to do such a computation.

All authors hold positions in the Department of Pediatrics, College of Physicians & Surgeons, Columbia University, New York, New York. Ralph B. Dell, M.D., is Emeritus Professor of Pediatrics, Steve Holleran, is Staff Associate, and Rajasekhar Ramakrishnan, Sc.D., is Director of the Biomathematics Division in Pediatrics.

Types of Experiments

Types of experiments include pilot and exploratory, those based on success or failure of a desired goal, and those intended to test a formal hypothesis. Each type is discussed briefly below.

Pilot and Exploratory Experiments

It is not possible to compute a sample size for certain types of experiments because prior information is lacking or because the success of the experiment is highly variable, such as in producing a transgenic animal. For other types of experiments, complicated statistical designs can be simplified to an important comparison wherein the sample size should be large enough to have a good chance of finding statistical significance (often called power; see Effect Size, Standard Deviation, Power, and Significance Level). Pilot experiments are designed to explore a new research area to determine whether variables are measurable with sufficient precision to be studied under different experimental conditions as well as to check the logistics of a proposed experiment. For example, suppose the investigator wishes to determine whether a certain factor, *x,* is elevated in an animal model of inflammation. The laboratory has developed an assay for factor *x* and now wishes to determine the variation of factor *x* in a population of mice. In the protocol, the investigator proposes measuring the concentration of factor *x* in 10 animals before and after the induction of inflammation. In a pilot experiment such as this, the number of animals to be used is based on experience and guesswork because there are no prior data to use in estimating the number of animals needed for the study. The experiment is performed to provide a rough idea of the standard deviation and the magnitude of the inflammatory effect.

A statistical analysis of the results yields estimates of the mean and standard deviation of factor *x* concentration before and after the induction of inflammation as well as estimates of the mean difference and its standard deviation. Such estimates can then be used to compute the sample size for further experiments. The investigator would be encouraged if the standard deviation of factor *x* in the 10 animals is relatively small compared with the concentration of the factor. Suppose that the mean concentration of factor *x* increased twofold after inflammation was induced, a change that should be easily detected if the variation of the change in the population is low. Then the pilot experiment will have

been encouraging in that the investigator may be able to track the increase in the concentration of factor *x* over time and determine changes in the concentration of the factor with various forms of therapy. The results of the pilot experiment can be used to estimate the number of animals needed to determine time trends and to study the effect of various interventions on the concentration of factor *x* using methods described below.

Sometimes "exploratory" experiments are performed to generate new hypotheses that can then be formally tested. In such experiments, the usual aim is to look for patterns of response, often using many different dependent variables (characters). Formal hypothesis testing and the generation of *p* values are relatively unimportant with this sort of experiment because the aim will be to verify by additional experiments any results that appear to be of interest. Usually the number of animals used in such experiments is based on a guess based on previous experience. Data collected in exploratory experiments can then be used in sample size calculations to compute the number of animals that will be needed to test attractive hypotheses generated by the exploration.

Experiments Based on Success or Failure of a Desired Goal

In experiments based on the success or failure of a desired goal, the number of animals required is difficult to estimate because the chance of success of the experimental procedure has considerable variability. Examples of this type of experiment are production of transgenic animals by gene insertion into fertilized eggs or embryonic stem cells. Large numbers of animals are typically required for several reasons. First, there is considerable variation in the proportion of successful gene or DNA incorporation into the cell's genome. Then there is variability in the implantation of the transferred cell. Finally, the DNA integrates randomly into the genome and the expression varies widely as a function of the integration site and transgene copy number.

Compounding this variability, different strains of mice react differently to these manipulations, and different genes vary in their rates of incorporation into the genome. It is often necessary to make several transgenic lines (see the discussion of transgenic animals in the ARENA/OLAW Institutional Care and Use Committee Guidebook [ARENA/OLAW 2002]). Using equation 1 below (Single-Group Experiments) and assuming that the success rate for all of the steps just mentioned is 5%, then one would need to use 50 animals, whereas a success rate of 1% would require using 300 animals. These numbers accord with the experience of investigators in the field and are usually the range of numbers of mice required to produce a single transgenic line.

In the case of knockout or knockin mice produced by homologous recombination, there is much less variability in the results and fewer animals may have to be produced. Again, it is difficult to predict the number required, especially if investigating the effects of regulatory sequences rather than of protein expression. The number of animals required is usually estimated by experience instead of by any formal statistical calculation, although the procedures will be terminated when enough transgenic mice have been produced. Formal experiments will, of course, be required for studying the characteristics of the transgenic animals requiring yet more animals.

Experiments to Test a Formal Hypothesis

Most animal experiments involve formal tests of hypotheses. In contrast to pilot experiments and the other types of experiments described above, it is possible to estimate the number of animals required for these experiments if a few items of information are available. Broadly, there are three types of variables that an investigator may measure: (1) **dichotomous variable,** often expressed as a rate or proportion of a yes/no outcome, such as occurrence of disease or survival at a given time; (2) **continuous variable,** such as the concentration of a substance in a body fluid or a physiological function such as blood flow rate or urine output; and (3) **time to occurrence** of an event, such as the appearance of disease or death. Many statistical models have been developed to test the significance of differences among means of these types of data. Detailed discussions of the models can be found in books on statistics (Cohen 1988; Fleiss 1981; Snedecor and Cochran 1989), in manuals for various computer programs used for statistical analyses (Kirkpatric and Feeney 2000; SAS 2000), and on websites that present elementary level courses on statistics (e.g., <http://www.ruf.rice.edu/~lane/rvls.html>). In this article, we describe methods for computing sample size for each of these types of variables.

Defining the Hypothesis to Be Tested

Although experimental designs can be complicated, the investigator's hypotheses can usually be reduced to one or a few important questions. It is possible then to compute a sample size that has a certain chance or probability of detecting (with statistical significance) an effect (or difference) the investigator has postulated. Simple methods are presented below for computing the sample size for each of the three types of variables listed above. Note that the smaller the size of the difference the investigator wishes to detect or the larger the population variability, the larger the sample size must be to detect a significant difference.

Effect Size, Standard Deviation, Power, and Significance Level

In general, three or four factors must be known or estimated to calculate sample size: (1) the effect size (usually the

difference between 2 groups); (2) the population standard deviation (for continuous data); (3) the desired power of the experiment to detect the postulated effect; and (4) the significance level. The first two factors are unique to the particular experiment whereas the last two are generally fixed by convention. The magnitude of the effect the investigator wishes to detect must be stated quantitatively, and an estimate of the population standard deviation of the variable of interest must be available from a pilot study, from data obtained via a previous experiment in the investigator's laboratory, or from the scientific literature. The method of statistical analysis, such as a two-sample *t*-test or a comparison of two proportions by a chi-squared test, is determined by the type of experimental design. Animals are assumed to be randomly assigned to the various test groups and maintained in the same environment to avoid bias. The power of an experiment is the probability that the effect will be detected. It is usually and arbitrarily set to 0.8 or 0.9 (i.e., the investigator seeks an 80 or 90% chance of finding statistical significance if the specified effect exists). Note that 1-power, symbolized as β, is the chance of obtaining a false-negative result (i.e., the experiment will fail to reject an untrue null hypothesis, or to detect the specified treatment effect).

The probability that a positive finding is due to chance alone is denoted as α, the significance level, and is usually chosen to be 0.05 or 0.01. In other words, the investigator wishes the chance of mistakenly designating a difference "significant" (when in fact there is no difference) to be no more than 5 or 1%. Once values for power and significance level are chosen and the statistical model (e.g., chi-squared, *t*-test, analysis of variance, linear regression) is selected, then sample size can be computed using the size of the effect the investigator wishes to detect and the estimate of the population standard deviation of the factor to be studied, using methods outlined below.

Several websites contain discussions of the principles of sample size calculations or have programs that will permit the user to make sample size calculations using various techniques. A few of these are

- <http://www.biomath.info>: a simple website of the biomathematics division of the Department of Pediatrics at the College of Physicians & Surgeons at Columbia University, which implements the equations and conditions discussed in this article;
- <http://davidmlane.com/hyperstat/power.html>: a clear and concise review of the basic principles of statistics, which includes a discussion of sample size calculations with links to sites where actual calculations can be performed;
- <http://www.stat.uiowa.edu/~rlenth/Power/index.html>: a site where sample size calculations can be made for many different statistical designs;
- <http://www.zoology.ubc.ca/~krebs/power.html>: a review of several software packages for performing sample size calculations; and
- <www.lal.org.uk/hbook14.htm> references an excellent handbook on experimental design and includes links to several statistical packages.

Also available are specialized computer programs such as nQuery Advisor, and statistical packages such as SPSS, MINITAB, and SAS, which will run on a desktop computer and can be used both for sample size calculations and for performing statistical analysis of data.

It should be noted that in the following discussion of sample size calculations, the aim is to simplify the question being addressed so that power calculations can be performed easily. There is no need to alter the actual design of the experiment and data analysis. Using, for example, randomized block, Latin square and/or factorial experimental designs, and the analysis of variance, it is possible to control for the effect of strain differences on a factor such as survival or response to an intervention and to obtain a more significant result than using more elementary methods. However, the simplified designs discussed here yield sample sizes close to what would be obtained with more complex analyses and hence should help the investigator be self-sufficient in planning experiments.

Experiments can be classified in a variety of ways. Many are carried out in two (or more) groups of animals. In the text below, these types are considered first, followed by single-group experiments.

Sample Size for Dichotomous Data

An experiment can involve measurement of dichotomous variables (i.e., occurrence of an event, expressed as rates or proportions). Sample size calculations for dichotomous variables do not require knowledge of any standard deviation. The aim of the experiment is typically to compare the proportions in two groups. In such a case, a relatively simple formula (Appendix Equation 1) will give the required sample size, given values for power, significance level, and the difference one wishes to detect. If more than two groups are studied, it is often possible to identify two rates that are more important to compare (or closest to each other) than any other pair.

Many books on statistics have tables that can be used to compute sample size, and nearly all statistical computer programs also yield sample size when power, significance level, and size of difference to be detected are entered. As an example, suppose previous data suggest that the spontaneous incidence of tumors in old rats of a particular strain is 20% and an experiment is to be set up to determine whether a chemical increases the incidence of tumors, using the same strain of rats. Suppose also that the scientist specifies that if the incidence increases to 50%, he/she would like to have an 80% chance of detecting this increase, testing at $p = 0.05$. Using Appendix Equation 1 and entering $p_1 = 0.2$, $p_2 = 0.5$ (for $power = 0.8$ and $\alpha = 0.05$), we learn that this experiment would require 43.2 or roughly 45 rats per group.

Note that the equations in the Appendix (also used in the calculations that can be carried out on the <www.biomath.info website) give sample sizes large enough to detect an increase or decrease in the variable (i.e., for a two-tailed test). Even when the postulated effect is an increase, it can be argued that a statistically significant change in the opposite direction is interesting and may merit further study. Nearly all clinical trials are now designed for two-tailed tests. In the carcinogenicity rat assay described above, it might be interesting and warrant further study if the test compound resulted in a significant fall in the spontaneous tumor rate. Also note that Appendix Equation 1 contains a continuity correction for the fact that the distribution of discrete data is being approximated by a continuous distribution (Fleiss 1981). Many computer programs used for sample size calculation do not include the continuity correction and hence will yield somewhat smaller sample size values.

Sample Size for Continuous Variables

Experiments are often designed to measure continuous variables such as concentration of a substance in a body fluid or blood flow rate. Although the statistical analytical models may be complex, it is often critical to detect the difference in the mean of a variable between two groups if that difference exists. In this case (Appendix Equation 2), a simple formula can be used to compute sample size when power, significance level, the size of the difference in means, and variability or standard deviation of the population means are specified. Again, the calculations are available in most modern statistical packages.

Suppose that in previous experiments the mean body weight of the rats used at a certain age is 400 g, with a standard deviation of 23 g, and that a chemical that reduces appetite is to be tested to learn whether it alters the body weight of the rats. Assume also that the scientist would like to be able to detect a 20 g reduction in body weight between control and treated rats with a power of 90% and a significance level of 5%, using a two-tailed unpaired *t*-test (two-tailed because the chemical might increase body weight). A computer program, or calculations based on Appendix Equation 2, suggests that 28.8 rats per group or roughly 60 (30 animals per group times 2 groups) rats are required for the whole experiment.

Single-Group Experiments

If the aim is to determine whether an event has occurred (e.g., whether a pathogen is present in a colony of animals), then the number of animals that need to be tested or produced is given by:

$$n = \frac{\log \beta}{\log p}, \qquad (1)$$

where 1-β is the chosen power (usually 0.10 or 0.05) and p represents the proportion of the animals in the colony that are not infected. Note that the proportion *not* infected is used in the formula. For example, if 30% of the animals are infected and the investigator wishes to have a 95% chance of detecting that infection, then the number of animals that need to be sampled (n) is

$$n = \frac{\log 0.05}{\log 0.7} = 8.4.$$

A total of nine animals should be examined to have a 95% chance of detecting an infection that has affected 30% of the animals in the colony. If the prevalence of infection is lower (e.g., 10%), then

$$n = \frac{\log 0.05}{\log 0.9} = 28.4.$$

Roughly 30 animals should be sampled. Thus, many more animals need to be sampled if the prevalence of the pathogen is low.

Proportion

The result described above is for a case in which the occurrence of an event in even one animal is of interest. In other single-group experiments, the researcher is interested in establishing that the postulated proportion is nonzero, or different from a prespecified value (known from prior studies, from physiological considerations, or as a value of clinical interest). It can be shown that the number of animals required for such an experiment is simply half the number given by Appendix Equation 1. In this case, p_e is the postulated proportion, and p_c is 0 or the prespecified value.

Continuous Variable

In a similar fashion, the researcher may measure a continuous variable in a single group and wish to establish that it is nonzero or different from a prespecified value. As with a proportion, it can be shown that the number of animals required for such an experiment is simply half the number given by Appendix Equation 2. In this case, d is the difference between the prespecified value and the postulated mean experimental value.

Controlling Variability by Repeat Study

Estimates of the required sample size depend on the variability of the population. The greater the variability, the larger the required sample size. One method of controlling for variability in the level of a continuous variable such as

blood flow is to measure the variable before and after an experimental intervention in a single animal. In this case, instead of using an estimate of the variability of the population mean, the variability of the difference is estimated. The standard deviation of the difference in a measurement in an individual is lower because it does not include interindividual variability. Stated in other terms, each animal is its own control. The number of animals needed to test a hypothesis will be reduced because the effect of animal-to-animal variation on the measurement is eliminated. Such an experiment is normally analyzed using a paired *t*-test. Appendix Equation 3 provides sample size calculation for such an experimental design. Crossover designs in which different groups of animals may have several different treatments in random sequential order are a generalization of this example. Such designs are also used to eliminate interindividual variability. In determining sample size, it is probably best to base the estimates on two chosen treatments.

Correlation Between Two Variables

If two continuous variables are measured in a single group, the question may be whether they are correlated significantly. For an assumed or postulated correlation coefficient, it is possible to calculate the number of animals needed to find a significant correlation. Appendix Equation 4 provides the necessary formula.

Considerations of Normality

The sample size calculations for continuous variables (Appendix Equations 2-4) assume that the variables are normally distributed (i.e., the values fall on a bell-shaped curve). The calculations are fairly robust: Small departures from normality do not unduly influence the test of the hypothesis. However, if the variable has a long tail in one direction (usually to the right), then the deviation from normality becomes important. A common method for making a distribution more normal is to use the log or square-root or some other transformation in the analyses. Such a transformation will often result in a variable that is closer to being normally distributed. One then uses the transformed variable for sample size calculations and for further statistical analysis.

Sample Size for Time to an Event

Simple Approaches

The statistical analysis of time to an event involves complicated statistical models; however, there are two simple approaches to estimating sample size for this type of variable. The first approach is to estimate sample size using the proportions in the two experimental groups exhibiting the event by a certain time. This method converts time to an event into a dichotomous variable, and sample size is estimated by Appendix Equation 1. This approach generally yields sizes that are somewhat larger than more precise calculations based on assumptions about the equation that describes the curve of outcome versus time.

The second approach is to treat time to occurrence as a continuous variable. This approach is applicable only if all animals are followed to event occurrence (e.g., until death or time to exhibit a disease such as cancer), but it cannot be used if some animals do not reach the event during the study. Time to event is a continuous variable, and sample size may be computed using Appendix Equation 2.

Unequal Number of Animals in Different Groups

Studies of transgenic mice often involve crossing heterozygous mice to produce homozygous and heterozygous littermates, which are then compared. Typically, there will be twice as many heterozygotes in a litter as homozygotes, although the proportions may be different in more complicated crosses. In such experiments, the researcher wishes to estimate the number of animals with the expected ratio between the experimental groups. The equations provided in the Appendix become considerably more complex. The reader is directed to our website for unequal sample size calculations (the expected ratio of group sizes is entered in place of the 1.0 provided on the chi-squared test on proportions web page): <http://www.biomath.info>.

Summary

In this article, we have discussed simple methods of estimating the number of animals needed for various types of variables and experiments. The thrust of the argument is that although analysis of the final set of data may involve sophisticated statistical models, sample size calculations can usually be performed using much simpler models. The aim of the calculation is to estimate the number of animals needed for a study, a value that is usually rounded up to yield an adequate number of animals for the study.

It is frequently true, in the authors' experience, that investigators err on the side of using too few animals rather than too many. This propensity results in a study that has too little power to detect a meaningful or biologically significant result. Roberts and colleagues (2002) did a meta-analysis of 44 animal experiments on fluid resuscitation and found that none of them had sufficient power to reliably detect a halving of death rate. To avoid this error, it is necessary to choose the power, the significance level, and the size of the effect to be detected, and to estimate the population variability of the variable being studied. Although the design of the experiment is simplified for the purposes of estimating sample size, it should be noted that using a more sophisticated design and statistical analysis usually yields the most power to detect any difference.

References

ARENA/OLAW [Applied Research Ethics National Association/Office of Laboratory Animal Welfare]. 2002. Institutional Animal Care and Use Committee Guidebook. Available from OLAW, National Institutes of Health, Rockledge 1, Suite 360, MSC 7982, 6705 Rockledge Drive, Bethesda, MD 20892-7982.

Cohen J. 1988. Statistical Power Analysis for the Behavioral Sciences. 2nd Ed. Mahwah NJ: Lawrence Erlbaum Associates.

Fleiss JL. 1981. Statistical Methods for Rates and Proportions. 2nd Ed. New York: Wiley.

Kirkpatric LA, Feeney BC. 2000. A Simple Guide to SPSS for Windows: Versions 8, 9, and 10. 4th Ed. Wadsworth Publishing Co. <www.wadsworth.com>.

Roberts I, Kwan I, Evans P, Haig S. 2002. Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation. Br Med J 324: 474-476.

SAS. 2000. Step-by-Step Programming with Base SAS Software. Cary NC: SAS Institute Inc.

Snedecor GW, Cochran WG. 1989. Statistical Methods. 8th Ed. Ames: Iowa State Press.

Appendix

Dichotomous Variables (Rates or Proportions)

Let r_c be the number of outcomes (an outcome is an event of interest such as occurrence of disease, death, or presence of a trait like coat color) in the control group, and r_e is the outcome in the experimental group.

Define

$$p_c = \frac{r_c}{N_c}; p_e = \frac{r_e}{N_e},$$

where r_c is the number of events and N_c is the total number of animals in control group or group c, and r_e, N_e for the experimental group or group e.

The investigator's hypothesis is that p_e is different from p_c. This hypothesis can be stated as a null hypothesis, H_0 (i.e., there is no difference between the two proportions), and a statistical test is devised to test that hypothesis. If the null hypothesis is rejected, then the investigator can conclude, at significance level α, that there is a difference between the two proportions. If the null hypothesis is not rejected, then the alternative hypothesis is rejected with the probability that a false-negative of β has occurred. These hypotheses can be stated as follows:

$$H_o: (p_c - p_e = 0)$$
$$H_1: (p_c - p_e \neq 0).$$

The formula for determining sample size is derived from a common statistical test for H_o. Usually the investigator knows or can estimate the proportion of the control group, which will have the outcome being observed, and can state a difference between the control group and the experimental group that he/she wishes to detect. The smaller this difference, the more animals will be needed. Thus, given estimates for p_c and p_e, sample size n for each group can be estimated:

$$n = C\frac{p_c q_c + p_e q_e}{d^2} + \frac{2}{d} + 2, \qquad \text{Equation 1 (Fleiss 1981)}$$

where $q_c = 1 - p_c$; $q_e = 1 - p_e$; and $d = | P_c - P_e |$. d is the difference between p_c and p_e, expressed as a positive quantity. C is a constant that depends on the values chosen for α and β. There is seldom justification for one-sided tests. The following list provides values of C for two levels of α and β for two-sided tests (i.e., detection of any significant difference if the experimental group is either higher or lower than the control group):

	α	0.05	0.01
1-β	0.8	7.85	11.68
	0.9	10.51	14.88.

If the observed $p_c = 0.5$ and the investigator wishes to detect a rate of 0.25 ($p_e = 0.25$), then $d = .25$. Further choose $\alpha = 0.05$ and $1\text{-}\beta = 0.9$ so $C = 10.51$. Then

$$n = 10.51\,\frac{0.5 \times 0.5 + 0.25 \times 0.75}{0.25^2} + \frac{2}{0.25} + 2 = 83.57$$

in each group, which when rounded off is 85 in each group for a total number of animals of 170.

Continuous Variables

Studies Comparing Two Group Means

To compute sample size for continuous variables, it is necessary to obtain an estimate of the population standard deviation of the variable (s) and the magnitude of the difference (d) the investigator wishes to detect, often called the effect. Sample size is given by

$$n = 1 + 2C\left(\frac{s}{d}\right)^2, \qquad \text{Equation 2 (Snedecor and Cochran 1989)}$$

where s is the standard deviation, d is the difference to be detected, and C is a constant dependent on the value of α and β selected. C can be determined from the table above, which gives values for C for two levels of α and β. Note that for $\alpha = 0.05$ and $1\text{-}\beta = 0.9$, C is 10.51 and 2C would be 21. If s is 4, d is 3, $\alpha = 0.05$, and $1\text{-}\beta = 0.9$ (i.e., $C = 10.51$ and $2C = 21$), then

$$n = 1 + 21\left(\frac{4}{3}\right)^2 = 38.37$$

in each group or roughly 80 animals for the whole study.

A useful rule of thumb is to multiply

$$\left(\frac{s}{d}\right)^2$$

(i.e., the quantity standard deviation divided by the difference to be detected squared) by 20 to obtain sample size for each group. For the example above, the rule of thumb ($20 \times \left(\frac{4}{3}\right)^2$) yields 35.5 or roughly 36 in each group.

Paired Studies

Paired studies compare values before and after an intervention in the same animal. In this case, data are analyzed by a paired *t* test, and the sample size is computed by

$$n = 2 + C\left(\frac{s}{d}\right)^2. \quad \text{Equation 3 (Snedecor and Cochran 1989)}$$

Note that

$$\left(\frac{s}{d}\right)^2$$

is multiplied by C in paired studies rather than 2C showing that paired studies are more powerful than comparison of two independent means.

Correlation Coefficient Between Two Continuous Variables in a Single Group

A correlation coefficient r (from n observations) does not have a normal distribution; however, the transformation

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

produces a normal approximation with standard error approximately $1/\sqrt{(n\text{-}3)}$ (Snedecor and Cochran 1989). From this calculation, the number of animals needed to show that a postulated (positive) correlation coefficient r is different from a specified r_0 is given by

$$n = 3 + \frac{4C}{\left[\ln\left(\frac{1+r}{1-r} \times \frac{1-r_0}{1+r_0}\right)\right]^2}, \quad \text{Equation 4}$$

where C is given in the list of C values above.

All four equations are implemented on our departmental web page. The web page also allows calculations of detectable effect size when the number of animals is given, in addition to allowing the number of animals to be different in the two study groups, as can happen in comparing heterozygous and homozygous littermates. As noted in the text, the link to the web page is <http://www.biomath.info>.