Ann Ist Super Sanità 2004;40(2):157-163

Biostatistical approaches to reducing the number of animals used in biomedical research

Maria PUOPOLO

Dipartimento di Biologia Cellulare e Neuroscienze, Istituto Superiore di Sanità, Rome, Italy

Summary - For ethical reasons, the least number of animals possible should be used in biomedical research, though not so few as to fail to detect biologically important effects or to necessitate the repetition of experiments. We describe biostatistical approaches that can contribute to either reducing the number of animals in single experiments or to increasing the quality of studies so that fewer subsequent studies (and thus animals) will be needed. The described approaches regard different phases of experimentation, specifically: planning the experimental design and calculating the sample size, controlling variability, choosing the response variable, postulating the statistical hypothesis to be tested, choosing the procedure for analysing data, and interpreting and suitably presenting the results.

Key words: reduction, experimental design, control group, experimental variability, response variable, statistical methods.

Riassunto *(Approcci biostatistici alla riduzione del numero degli animali utilizzati nelle sperimentazioni biomediche)*. - Per ragioni etiche, nelle sperimentazioni biomediche è opportuno utilizzare il minor numero possibile di animali, ma non una quantità così bassa da impedire l'identificazione di effetti biologicamente importanti o da richiedere la ripetizione degli esperimenti. In questo articolo vengono descritti gli approcci biostatistici che possono contribuire alla riduzione della numerosità degli animali nelle singole sperimentazioni o all'incremento della qualità intrinseca delle sperimentazioni stesse, in modo che successivamente siano necessari meno studi (e, conseguentemente, meno animali). Gli approcci qui discussi riguardano diverse fasi dell'esecuzione di un esperimento, e più specificamente la pianificazione del disegno sperimentale, il calcolo della numerosità campionaria, il controllo della variabilità, la scelta della variabile risposta, la postulazione dell'ipotesi statistica da verificare, la selezione della procedura per l'analisi dei dati e, infine, l'interpretazione dei risultati e la loro presentazione in forma adeguata.

Parole chiave: riduzione, disegno sperimentale, gruppo di controllo, variabilità sperimentale, variabile risposta, metodi statistici.

Statistical approaches to the 3R rule of animal experimentation

Since their development by W.M.S. Russel and R.L. Burch in 1959 [1], the 3R rule of the humane treatment of animals in research (i.e., Replacement, Reduction, and Refinement) have contributed to improving both the wellbeing of experimental animals and the quality of experiments. Regarding statistical approaches to the 3R, these have mainly contributed to Reduction, although when reducing the number of animals, the data often do not suffice for performing a meaningful statistical analysis and the experiment must be repeated. Statistical approaches may also contribute to Refinement by allowing more information to be obtained without increasing the number of animals, although practically no studies have adopted this approach. Recently, statistical knowledge has also contributed to Replacement, with computer and mathematical modelling being used as alternatives to animal experimentation.

Whatever the approach, for both ethical and economic reasons, it is important to use the least number of animals possible, though not so few as to fail to detect biologically important effects or to necessitate the repetition of experiments [2]. In general, the number of animals should be reduced without compromising the scientific quality of research or disrupting scientific progress. In reviews of the literature, it has been suggested that the number of laboratory animals used could often have been reduced while still obtaining statistically valid data [3-5]. With regard to animal use in Europe, in 1997, during the "Target 2000" Conference held in Brussels, it was

Indirizzo per la corrispondenza (Address for correspondence): Maria Puopolo, Dipartimento di Biologia Cellulare e Neuroscienze, Istituto Superiore di Sanità, V.le Regina Elena 299, 00161 Rome, Italy. E-mail: mpuopolo@iss.it.

stated that an adequate biostatistical approach could contribute to reducing the number of animals used in biomedical research and testing in Europe by 50%. In 1998, FRAME (Fund for the Replacement of Animals in Medical Experiments) established a "Reduction Committee" consisting of experienced professionals in the fields of statistics, experimental design, and animal welfare, among others, with the aim of promoting suitable experimental designs and statistical methods.

The use of animals in research is in large part justified by the results obtained; thus, from an ethical standpoint, communicating these results is crucial, even when they are not statistically significant. In this way it can be ensured that duplicate studies that generate redundant data are not carried out, which would decrease the use of animals. In fact, the widespread communication of results is a "benefit" factor in many models of the ethical assessment of animal research [6]. It is thus the researcher's responsibility to publish the results as accurately and completely as possible and to the widest possible audience.

In the present work, we describe biostatistical approaches that can contribute to reducing the number of animals used in research, whether by reducing the number of animals necessary for obtaining the desired information in a single study or by optimising the quality of studies so that fewer subsequent studies will be needed to understand the phenomenon being investigated. The described approaches regard, and are important for, different phases of experimentation, specifically: planning the experimental design and calculating the sample size, controlling variability, choosing the response variable, postulating the statistical hypothesis to be tested, choosing the procedure for analysing data, and, last but not least, interpreting and suitably presenting the results [3].

The optimal experimental design

A well-designed study ideally allows biases to be avoided and is sufficiently powerful to produce statistically significant results, which can allow the repetition of studies to be avoided. In general, the design should conform to one of the formal designs described in the literature [7-11], or to a combination of these designs, each of which considers the specific features and constraints of a given experimental material and of the nature of the investigation. A brief description of several types of study designs is provided below.

The simplest design is the completely randomized design (CRD), which assumes that independent experimental units (e.g., laboratory animals) are assigned to different levels of a factor under study (e.g., different treatment groups). To guarantee that the CRD is properly implemented, the different units must be assigned to the different levels of the factor in a random manner, meaning that every unit (or subject) has an equal chance of being assigned to any of the experimental levels (or groups). Randomization ensures that the response of one or more of the groups is not biased by the effects of unknown variables or by sources of variation that cannot be controlled. For example, among animals housed in the same cage, being among the first animals to have been captured could have an effect on the response being studied; thus randomization should be used so that these animals are not assigned to the same group. The ultimate goal is to equalise the groups. Although this cannot be guaranteed, the chance of creating unequal groups diminishes as the sample size increases. To facilitate the randomization process, tables of random numbers or specific computer software can be used.

In the completely randomized factorial design (CRFD), the experimental units are randomly assigned to combinations of levels of two or more factors, and CRFD allows the effects of the different factors and their interactions to be evaluated simultaneously, although this entails a heavier workload for the researcher. Obviously, in factorial designs, all of the levels of a given factor must be constant in the different levels of the other factors (e.g., when different drugs are evaluated on different days following treatment, the levels of the factor "time elapsed from treatment to evaluation" must be the same for the different drug groups). These designs are particularly useful in the pharmaceutical industry, where compounds are subjected to a cascade of screening tests for assessing biological activity against a specific target. Many factors can influence the response of experimental animals to treatment (e.g., gender, strain, age, and protocol-specific factors, such as the timing and means of administering treatment). CRFD can be used to explore which factors and levels of these factors will maximise the difference between a control and a treatment, which is more effective and efficient than varying one factor at a time. The results can be used to design more efficient experiments, either by reducing the number of animals or by increasing the sensitivity of the experiment, so that smaller biological effects can be detected. This could help to reduce animal use, especially in the pharmaceutical industry [12].

In a randomized block design (RBD), all of the different levels of a factor are randomly assigned to experimental units of the same block. A block is a group of units characterised by dependency (e.g., pups of the same litter, individuals of the same family, animals

housed in groups, or members of a social group), which renders the units of the same block homogeneous and the units of different blocks heterogeneous. A block factor should also be recognised when an experiment is divided into two or more parts carried out at different times or by different researchers. In RBD, by controlling variability among blocks, the effect of a factor is evaluated with greater precision and can thus be highlighted using fewer experimental units. The relative efficiency [13] of RBD, with respect to CRD, is used to determine whether or not the reduction in residual variability, and thus the increase in the precision of a RBD, justifies the complexity of taking into account the blocks (both for designing and conducting the experiment and for analysing the data). Examples of RBD include repeated-measure designs, in which all of the levels of an experimental factor (e.g., different stimuli, drugs, or drug doses) are tested on the same subject in a random order with wash-out intervals, and growth-curve designs.

A full factorial design entails evaluating all combinations of treatment and can thus require numerous subjects. However, if the principal objective of an experiment is to estimate and test the main effects and a few lower-order interactions and if it can be safely assumed that the other effects are negligible, an incomplete block design or a fractional factorial design can provide the desired information, using much fewer animals. In incomplete block design, only some of the combinations of treatment are assigned to each block, in such a way that across the blocks each treatment combination is evaluated the same number of times. Analogously, fractional factorial design requires testing only a fraction of the treatment combinations that would be needed for a full factorial design. When adopting an incomplete block design or a fractional factorial design, the researcher must specify the effects that he/she wants to include in the model so as not to have more parameters than it is possible to estimate. Specifically, it is necessary to exclude the negligible effects of the interactions between the block and treatment or of the interactions among treatments.

A Latin square design (LSD) allows the main effects of three different factors, each with k levels, to be evaluated using only k·k subjects (the corresponding full factorial design would require k·k·k subjects). In an LSD, k animals undergoing one of the k levels of a factor and each animal undergoing a different level of the second factor are randomly assigned to k levels of the third factor. This assigning of animals to the combination of the levels of the three factors is repeated for all levels of the first factor, taking care that the combinations of the levels of the second factor with the levels of the third factor are different in the different levels of the first factor. LSD is used when the interactions are negligible, usually to balance nuisance factors such as the order in which the drugs are administered or previous treatments, including also different degrees of manipulation. When a fourth factor emerges, which can simply consist of the researcher, an LSD may be modified into a Greco-Latin square design (GLSD) by assigning the levels of the fourth factor as done for the third factor, ensuring that none of the combinations between levels of the third and fourth factors repeat in the same square. The factors involved in an LSD or in a GLSD may differ; they may be both fixed effect factors and/or random effect factors, and it is crucial to consider their nature in the analysis, so as to increase efficiency. To increase the degree of freedom, the squares may be repeated. The schemes of LSD and GLSD for some values of k are reported in the textbook of Fleiss [13].

In split plot design (SPD), the levels of one factor (so-called "subplot treatments") are applied to the ultimate, single experimental units, and the levels of the other factor (the so-called "whole-plot treatment") are applied to sets of the ultimate units. SPD is frequently used in studies of the interactions of prenatal treatments with postnatal treatments, where pregnant subjects are exposed to the different levels of a factor (or a combination of levels of factors) and offspring are subsequently exposed to different levels of a factor (or a combination of levels of factors).

In sequential experimental designs (SED), the decision as to the total number of animals to test is reviewed as each animal's data is collected and evaluated. SED are more powerful than other designs, in that they require fewer subjects to come to a conclusion with the same degree of certainty. However, these designs are neither suitable for rare or long-lasting events, nor when the experiments require long periods of time or are too expensive to be carried out for a single experimental subject at a time. Nevertheless, SED are the most ethical alternative when attempting to reduce the number of animals to the absolute minimum, which is especially desirable when experiments are particularly aversive [14].

In planning experiments, pilot studies can provide useful information and are commonly used for new procedures or when new compounds are being tested. Pilot studies use a small number of animals to generate preliminary data which provide evidence supporting the rationale of the proposal; they can also be used to further evaluate the study question, hypotheses, and procedures.

When possible, experiments should be conducted "blindly" with respect to the treatment, using coded samples, so that the treatment group remains unknown until the data are analysed.

Choosing the control group

The use of a control group is important for minimising the impact of extraneous variables or for recognising unwanted variables. In positive controls, changes are expected and act as a standard against which to measure differences in severity among experimental groups. Experimental groups are investigated to determine whether these alterations may be prevented or cured. Negative controls are expected to produce no change from the normal state and are used to ensure that an unknown variable is not adversely affecting the animals, which might result in a false-positive conclusion. Sham controls are used to mimic a procedure or treatment without actually performing the procedure or using the substance being tested. For example, in the surgical implantation of "X" in the abdominal cavity, the sham control undergoes the same surgical procedure as the treated animal (i.e., the abdominal cavity is opened) yet without having the "X" implanted. Another example of a sham control is the use of a placebo in drug trials. In vehicle controls, the supposedly innocuous substance (e.g., saline or mineral oil) acting as a vehicle for the experimental compound is administered alone and in the same manner in which it will be used with the experimental compound, making it possible to determine whether the vehicle itself has any effects. A comparative control often consists of a positive control that receives a clinically accepted treatment and is used for direct comparisons with an animal receiving a new treatment. In certain cases, the same control group can be used for more than one treatment group. The use of historical controls can also be considered, especially for endpoints and control substances that have been commonly used in animal experimentation and that have provided consistently concordant results.

Controlling experimental variability

In designing experiments, potential variability and its effect on the results should be taken into account, so that the experiment will be efficient and provide the most reliable information, ultimately contributing to reducing the use of animals.

Many variables, in addition to those related to the experiment itself, can affect biological responses and thus produce biases in the results, especially when such variables are not constant for the different experimental groups [15, 16]. Variability can be controlled by: a) using homogeneous animals (inbred; with the same previous experience or type of manipulation; same age and/or weight); b) excluding the presence of diseases, including both infectious diseases and parasites, with particular regard for sub-clinical conditions; c) controlling environmental variables and/or the quality of husbandry (ventilation, temperature, light cycles, relative humidity, room noise levels, ammonia levels, cage type, population density, and quality of food and water), in addition to smell hormones produced by animals during the actual experiment, which can frighten or cause stress to the other animals; d) adopting the same standardised methodology; and e) using a suitable method for accurately measuring the response variable.

All of these factors should be considered in the designing and planning of the experiment. However, in experiments in which this has not been done, or to confirm the validity of these factors, some confounders and/or biases may be evaluated by including blocks or covariates in the statistical analysis. During the actual experiment, in addition to the primary response variable, many variables are often recorded and may be incorporated into the statistical analysis so as to increase precision and power (e.g., age, body weight at several points in time, food and water consumption, haematology, clinical biochemistry, and behavioural response). These variables can also be useful in assessing the randomisation and in identifying outliers. They can also contribute to new hypotheses being formulated, to increasing the potential to generalise the results, and to explaining differences in results among different experiments or studies, not to mention that the appropriate use of these variables could increase the knowledge in a given area, reducing the number of experiments and thus the number of animals used.

Choosing the response variable

The number of subjects to be used depends upon the nature of the response variable. For given values of both type I error probability and statistical power, a binary response (dichotomous variable) requires the largest sample, with smaller samples needed for a categorical variable. A continuous variable requires the smallest sample and is the most powerful in detecting differences; thus, when possible, continuous variables should be used.

Formulating the statistical hypothesis

Clearly defining the problem statement, objectives, and hypotheses is crucial to the success of an experiment. Firstly, the specific questions of a working hypothesis and biologically significant effect must be addressed. The hypothesis should then be expressed in two distinct and clearly defined outcomes: a null

hypothesis, defined as no difference among experimental groups, and an alternative hypothesis, defined as a true difference among experimental groups.

Clearly identifying objectives that allow more precise hypotheses to be developed may reduce the number of experimental subjects. In certain cases, a general problem statement may be subdivided into multiple precise objectives. For example, the problem "to determine the time of a distinct peak" can be evaluated separately from the problem "to determine the size of the peak", so as to avoid using the number of animals that would be needed to achieve the latter for each point in time at which the peak occurs [17].

Choosing the most appropriate statistical analysis

The objective of a statistical analysis is to extract the most information possible from the data and to take into account biological variability and measurement error in interpreting the data. Statistical analyses are usually performed using standardised methods, the choice of which depends on a variety of factors. For example, to analyse numerical data, analysis of variance (ANOVA) can be used [7-11, 18-20]. In certain cases, data need to be transformed to satisfy assumptions, though most parametric methods are robust against moderate departure from the assumptions. A log transformation is often appropriate when the dependent variable is a concentration, whereas when data are expressed as proportions or percentages a logit or angular transformation may be suitable. Counts with low means (sampled from a Poisson distribution) may be transformed by the square root of the observations, whereas transformation may not be necessary when the mean counts are reasonably high. When the assumptions of ANOVA are violated, non-parametric methods, which entail replacing individual observations with their ranks, can be used [19, 20]. However, these methods result in the loss of some information and thus lack power in situations where parametric tests would be more suitable, although they may be more powerful in cases where parametric methods are not appropriate.

In addition to choosing between parametric and non-parametric tests or between transformed and non-transformed data, the choice of the most appropriate statistical test and/or data-analysis technique depends on a number of factors. For example, for censored data, a survival analysis is preferable to a simple t-test; for factorial designs, ANOVA is more appropriate than single t-tests; and to control for the effect of covariates that could be responsible for the observed response, covariance analysis should be used instead of separate analyses for the response variable and the covariates. Moreover, blocks or other random-effect factors should be included in the model; post-hoc comparisons or correction criteria to control for the probability of type I errors should be adopted [7-11, 19-21]; and differences in variability among groups should be evaluated. For categorical or qualitative data, the chi-square test or Fisher's exact probability test should be used, whereas for ordinal categorical data the chi-square test for trend is more appropriate than the simple chi-square test. In certain cases it may be more appropriate to perform an analysis of the correlation between two numerical variables (or between their ranks) or a regression analysis to quantify the relationship between a dependent variable and one or more independent variables.

The most appropriate statistical procedure should be chosen by an expert. Moreover, a deeper understanding of the phenomenon being investigated, and a consequent reduction in the number of experiments (and thus in the number of animals used) could be obtained through ad hoc analyses for complex structured data or multivariate statistical analyses [3, 22]. Previous knowledge of the investigated phenomena can be exploited using a Bayesian procedure [23], although this requires a complex formulation of the problem and intense calculations, implemented in specific software that is not easily accessible. For these reasons, the Bayesian approach can only be carried out by a qualified statistician, which probably explains why it is so rarely adopted. When similar studies have been previously conducted, their results, if applicable and valid, may be included in a meta-analysis [24], which allows the given hypotheses to be evaluated without using additional experimental subjects.

Data-quality control

The researcher has the moral responsibility of properly performing data collection, entry, and analysis. To detect errors in data entry, quality assurance procedures should be developed and incorporated into the experimental design (graphic methods are available in most statistical packages). If outliers are identified, the accuracy of the records should first be checked, so as to exclude data-entry errors; it should also be confirmed that there were no unusual occurrences or circumstances in performing the experiment on the animal in question. If no errors or unusual circumstances are discovered, the outlier can be discarded, although it could be useful to perform the statistical analysis both with and without the suspected data, so as to evaluate the effect on the results. When multiple outliers are present, it should be

checked whether or not the experiment had been performed sloppily, and statistical analysis should be used to determine whether these were true outliers or instead actual outcomes of the research which should not be discarded.

Interpreting and presenting results

The amount and quality of information drawn from the obtained data depends on the researcher's ability to interpret and present the data, which derives from both personal experience and knowledge of the theory of biostatistical methods. Researchers should avoid making claims about the effects of treatment when the results of a study are statistically significant but of little or no biological importance. The magnitude of statistically significant effects should always be specified with a measure of precision (confidence interval, standard deviation, or standard error) and exact p-values, and the biological relevance of the effects should be addressed.

When a non-statistically significant effect is found, researchers should use caution in concluding that the treatment had no effect, in that the result could be due to an insufficiently large sample size or to great experimental variability. A power analysis should be used to show the magnitude of the biological effect that the study was probably able to detect.

The descriptive statistics chosen to synthesise the data obtained should be consistent with the criteria adopted for the statistical analysis. For example, when a non-parametric approach is used, medians and interquartile ranges should be reported instead of means and standard deviations. When presenting the results of multiple comparisons, pooled standard error, obtained from the mean square error of the ANOVA results, should be reported, rather than single standard errors (or standard deviations). When interactions reach statistical significance, the results and comments regarding the main effects should be carefully considered. Finally, percentage frequency should be calculated for categorical data.

Calculating sample size and statistical power

There is no justification for failing to estimate the appropriate sample size or statistical power of an experiment [25, 26]. In fact, a mathematical relationship exists among the size of the effect of interest (the difference of biological significance), the standard deviation (usually taken from previous studies), the significance level (usually 0.05 or 0.01), the statistical power (usually 80% or 90%), the type of alternative hypothesis (bi-directional or unidirectional), and the sample size; and any five of these parameters can be used to determine the remaining parameter. Although the formulae are complex, especially when the statistical problem is not a mere comparison between the means of two experimental groups, software packages are available for carrying out these calculations [27, 28]. In planning an experiment, the unknown parameter is the sample size, that is, the number of subjects needed to produce a definite result.

Conclusions

Refining statistical approaches so that the data collected and analysed are as accurate, valid, and informative as possible can contribute to reducing the use of animals in research. Not only can these approaches reduce the number of animals necessary for a single experiment but they can also, and perhaps more importantly, increase the quality of studies, optimising the productivity of the animals and allowing the repetition of experiments to be avoided. Well designed and analysed experiments generate scientifically valid and reproducible data, which should be the ultimate goal of any scientific investigation, and the use of poorly designed or analysed animal experiments is no longer justifiable from an ethical standpoint.

Acknowledgements

This study was supported by the Italian Ministry of Health (Project no. 01/C "Biostatistical and ethological approaches for the promotion of welfare of laboratory animals and of the quality of experimental data", to Maria Puopolo).

Submitted on invitation.
Accepted on 3 March 2004.

REFERENCES

1. Russel WMS, Burch RL. *The principle of humane experimental technique*. London: Methuen & Co. Ltd; 1959.
2. Festing MF, Altman DG. Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J* 2002;43(4):244-58.
3. Still AW. On the number of subjects used in animal behaviour experiments. *Anim Behav* 1982;30:873-80.
4. McConway K. The number of subjects in animal behaviour experiments: is Still still right? In: Stamp Dawkins M, Gosling LM (Ed.). *Ethics in research on animal behaviour*. London: Academic Press; 1992. p. 35-8.
5. Festing MFW. Reduction of animal use: experimental design and quality of experiments. *Lab Anim* 1994;28:212-21.

6. Christiansen SB, Sandoe P. Ethics in animal behaviour and welfare research. In: *Proceedings of the 34th International Congress ISAE*. Florianopolis, 17-20 October. Brazil; 2000. p. 88.

7. Edwards AL. *Experimental design in psychological research*. New York: Harper & Row; 1985.

8. Wilcox RR. *New statistical procedures for the social sciences. Modern solutions to basic problems*. New Jersey: Lawrence Erlbaum, Hillsdale; 1987.

9. Cobb GW. *Introduction to design and analysis of experiments*. New York: Springer; 1998.

10. Cox DR, Reid N. The theory of the design of experiments. In: *Monographs on Statistics and Applied Probability, 86*. Boca Raton: Chapman & Hall/CRC Press; 2000.

11. Dean AM, Voss D. *Design and analysis of experiments*. New York: Springer; 1999.

12. Shaw R, Festing MF, Peers I, Furlong L. Use of factorial designs to optimize animal experiments and reduce animal use. *ILAR J* 2002;43(4):223-32.

13. Fleiss JL. *The design and analysis of clinical experiments. Wiley series in probability and mathematical statistics*. New York: John Wiley & Sons; 1986.

14. Chamove AS. Using fewer research animals. *Lab Prim Newsletter* 2003;42(1):1-2.

15. Festing MFW. Experimental design and husbandry. *Exp Gerontol* 1997;32:39-47.

16. Mering S, Kaliste-Korhonen E, Nevalainen T. Estimates of appropriate number of rats: interaction with housing environment. *Lab Anim* 2001;35(1):80-90.

17. Fry D. Reduction by well-defined objectives. In: *Proceedings of the 4th World Congress on alternatives and animal use in the life sciences*. New Orleans, 11-15 August 2002. (2004, in press).

18. Festing MFW. Guidelines for the design and statistical analysis of experiments in papers submitted to ATLA. *Alter Lab Anim* 2001;29:427-46.

19. Marascuilo LA, McSweeney M. *Nonparametric and distribution-free methods for the social sciences*. Monterey (Ca): Brooks/Cole; 1977.

20. Siegel S, Castellan NJ.Jr. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill; 1988.

21. Ludbrook J. Multiple comparison procedures updated. *Clin Exp Pharmacol Physiol* 1998;25(12):1032-7.

22. Puopolo M, Chiarotti F. Analysis of behavioral sequences: refinement of statistical methods to reduce animal use in the life sciences. In: Balls M, van Zeller AM, Halder M (Ed.). *Progress in the reduction, refinement and replacement of animal experimentation*. Amsterdam: Elsevier; 2000. p. 773-8.

23. Berger JO. *Statistical decision theory and bayesian analysis. Springer series in Statistics*. New York: Springer-Verlag; 1993.

24. Sutton AJ, Abrans KR,. Jones DR, Sheldon TA, Song F. *Methods for meta-analysis in medical research*. Chichester: John Wiley & Sons; 2000.

25. Thomas L, Juanes F. The importance of statistical power analysis: an example from animal behaviour. *Anim Behav* 1996;52:856-9.

26. Schwagmeyer PL, Mock DW. How to minimize sample sizes while preserving statistical power. *Anim Behav* 1997;54:470-4.

27. Cohen J. *Statistical power analysis for the behavioural science*. Hillsdale: Lawrence Erlbaum; 1988.

28. Erdfelder E, Faul F, Buchner A. GPOWER: a general power analysis program. *Behav Res Methods Instrum Comput* 1996;28(1):1-11.