Reducing the Use of Laboratory Animals in Biomedical Research: Problems and Possible Solutions

The Report and Recommendations of ECVAM Workshop 29[1,2,3]

Michael F.W. Festing,[4] Vera Baumans,[5] Robert D. Combes,[6] Marlies Halder,[7] Coenraad F.M. Hendriksen,[8] Bryan R. Howard,[9] David P. Lovell,[10] Graham J. Moore,[11] Philip Overend[12] and Marie S. Wilson[13]

[4]MRC Toxicology Unit, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK; [5]Department of Laboratory Animal Science, University of Utrecht, 3508 TD Utrecht, The Netherlands; [6]FRAME, Russell & Burch House, 96–98 North Sherwood Street, Nottingham NG1 4EE, UK; [7]ECVAM, JRC Environment Institute, 21020 Ispra (VA), Italy; [8]National Institute of Public Health and the Environment, 3720 BA Bilthoven, The Netherlands; [9]Field Laboratories, Medical School, University of Sheffield, Beech Hill Road, Sheffield S10 2RX, UK; [10]BIBRA International, Woodmansterne Road, Carshalton, Surrey SM5 4DS, UK; [11]Pfizer Central Research, Sandwich, Kent CT13 9NJ, UK; [12]SmithKline Beecham Pharmaceuticals, New Frontiers Science Park (North), Third Avenue, Harlow, Essex CM19 5AW, UK; [13]Merck Sharp & Dohme Research Laboratories, Neuroscience Research Centre, Terlings Park, Harlow, Essex CM20 2QR, UK

Preface

This is the report of the twenty-ninth of a series of workshops organised by the European Centre for the Validation of Alternative Methods (ECVAM). ECVAM's main goal, as defined in 1993 by its Scientific Advisory Committee, is to promote the scientific and regulatory acceptance of alternative methods which are of importance to the biosciences and which reduce, refine or replace the use of laboratory animals. One of the first priorities set by ECVAM was the implementation of procedures which would enable it to become well-informed about the state-of-the-art of non-animal test development and validation, and the potential for the possible incorporation of alternative tests into regulatory procedures. It was decided that this would be best achieved by the organisation of ECVAM workshops on specific topics, at which small groups of invited experts would review the current status of various types of *in vitro* tests and their potential uses, and make recommendations about the best ways forward (1). In addition, other topics relevant to the Three Rs concept of alternatives to animal experimentation have been considered in several ECVAM workshops. This is a report of the first ECVAM workshop to be devoted exclusively to *reduction* as defined by Russell & Burch (2).

The workshop on Reducing the Use of Laboratory Animals in Biomedical Research:

Address for correspondence: Dr Michael F.W. Festing, MRC Toxicology Unit, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK.

Address for reprints: ECVAM, TP 580, JRC Environment Institute, 21020 Ispra (VA), Italy.

[1]ECVAM — European Centre for the Validation of Alternative Methods. [2]This document represents the agreed report of the participants as individual scientists. [3]Derek Fry (Home Office, Shrewsbury, UK) was also a participant at the workshop and contributed to this report.

Problems and Possible Solutions was held in Southwell, UK, on 12–15 January 1998, under the chairmanship of Michael Festing (MRC Toxicology Unit, Leicester, UK). The participants, who all attended as individuals, not as representatives of their respective organisations, were very experienced in the use of animals in biomedical research, having a strong commitment to high quality research and the ethical use of animals where such use cannot be avoided. The aims of the workshop were to find ways of reducing the number of animals used in biomedical research without reducing research output, and to make recommendations for practical ways in which this might be achieved.

Introduction

The concept of the Three Rs (*replacement*, *reduction* and *refinement*) was developed by Russell & Burch (2) to provide a framework for improving the conduct and ethical acceptability of experimental techniques on animals. Given that animals used in research may experience pain, suffering or lasting harm, the first step must be to consider whether less sentient or non-sentient alternatives can be used instead (*replacement*). Where this is not possible, care needs to be taken to minimise any pain that an individual animal may suffer (*refinement*), both during the actual experiment and before or after the conduct of experiments. Refinement is often achieved, for example, by providing the animals with an environment in which they can feel secure and comfortable, ensuring that they are free of infectious diseases, and by using appropriate anaesthetics and analgesics if surgical techniques are to be used (3). Lastly, the number of animals used in a given project needs to be minimised (reduction), while ensuring that the objectives of the study can still be achieved; typically, this will also reduce the sum total of animal suffering.

For the purposes of this report, reduction means ways of obtaining comparable levels of information from the use of fewer experimental animals, or of obtaining more information from a given number of animals, so that fewer animals are needed to complete a given research project, taking into account individual animal welfare in relation to minimising pain, suffering, distress or lasting harm.

Although the concept of reduction is relatively simple, possible methods of achieving it are not immediately obvious. However, there is often a clear association between reduction and the quality of the resulting science. If animals are used in a poor quality project, which does not make a significant contribution to knowledge, or if a project is undertaken in such a way that it fails to meet its scientific objectives, animals will have been used needlessly. This will also be the case if the scientific objectives are not clear, so that it is not easy to determine whether these have actually been met. Even when the project is of high scientific calibre, there might be scope for reducing animal use by using pilot experiments and/or more advanced experimental designs and statistical methods. Very often this will also reduce the need for other scientific resources and avoid unnecessary work, so it will improve general scientific effectiveness in the long term.

One obvious area where there could be scope for reducing animal use is in ensuring that each experiment is of an appropriate size. Although a large experiment will usually have higher statistical power (i.e. it will be more likely to detect a treatment effect if there is one) than a comparable smaller one, it could lead to inefficient use of resources (including animals), because once an experiment has reached a certain size, the use of additional animals provides relatively little further information (4). However, animals will also be used unnecessarily if an experiment is so small that it is incapable of detecting a scientifically important treatment effect.

The exact design of an experiment can also be important. For example, randomised block experimental designs sometimes help to remove variability due to otherwise uncontrollable time and space variables, and they therefore increase the statistical power, so that fewer animals are needed. Factorial designs can be used to incorporate both sexes and/or more than one strain of animal without increasing the total number. If the treatment is eventually to be studied in both sexes, then such designs can reduce the numbers of animals needed and produce information which could not be obtained in any other way. Thus, reduction needs to be considered

in terms of research strategy, including the actual design, size and scientific information provided by each experiment.

The potential conflict between reduction and refinement should also be considered. It is sometimes possible to reduce the number of animals needed by increasing, for example, the dose of a test chemical to ensure that a toxic effect is observed. This may lead to the use of fewer animals, although each individual animal may suffer more. As it is difficult to quantify pain and suffering, care will need to be taken in such cases to ensure that proposed changes are not counter-productive in terms of total suffering. However, surveys suggest that there are many experiments which could be conducted with fewer animals without an increased burden on those which have to be used (3, 4), and that, in some cases, the experiments could be redesigned and analysed more efficiently to provide more information. Where increased refinement (for example, through the use of analgesics or by environmental enrichment) also reduces variation between animals, it may contribute to reduction. Thus, with careful thought at the design stage, fewer animals can often be used without any loss of information.

Reduction can also be achieved by minimising the wastage of animals which have been bred for research, but which are not used because of failure to match supply and demand. For example, demand may be largely for males so that females are not needed, or it may be so sporadic that a given batch of animals may have become too old or too heavy by the time they are required. Although these animals are not used for research, excess production still poses an ethical problem because the breeding and killing of animals for no real purpose is itself ethically undesirable, and their availability could encourage researchers to use more animals than they would otherwise consider necessary. Matching breeding to realistic use is particularly important for colonies of harmful mutant or transgenic animals, where at least some of the offspring may suffer adverse effects. Cryopreservation can reduce the need to maintain colonies simply to preserve the line.

Appropriate designs can also help to reduce such wastage. Factorial designs can often be used to even out the demand for both sexes, because splitting a single-sex group into two half-sized groups of males and females presents no serious statistical problems, and has the advantage of being able to show whether the two sexes respond in the same way. Similarly, animals heterogeneous for weight or age can be incorporated into an experiment without any loss of precision by using a randomised block design, obviating the need for very narrow weight or age ranges. Some of these points are considered in more detail later in this report.

Russell & Burch (2) suggested that reduction can be achieved by better research strategy, by better control of variation and by the application of better statistical methods.

Improving Research Strategy

There are several ways in which research strategy can be improved, as discussed below.

Objectives

Research objectives need to be clearly specified and flexible, with the definition of appropriate decision points. The latter would help a researcher to decide whether to continue with a particular line of research or to try another approach.

The most appropriate animal model should be chosen. A wide range of inbred strains, mutants, outbred stocks and transgenic strains of mice and rats are available, and the outcome of the project may depend critically on the strain(s) used. The choice of individual strains or stocks needs to be given careful consideration, and should be justified in research proposals. It has been claimed that "the introduction of inbred strains into biology is comparable in importance with that of the analytical balance into chemistry" (5). The uniformity of inbred strains means that, in many cases, fewer animals are needed than if outbred stocks are used (6), and selection of the most appropriate inbred strain from those which are available may lead to further reduction (7). If outbred stocks have normally been used in the past, the possibility of switching to inbred strains should be considered as a way of improving the science as well as of reducing animal numbers (8). However, whether inbred strains or outbred stocks are used, research workers should make some attempt to justify their

choice to indicate that they have at least given it some thought.

Background research

In evaluating the need to undertake a particular project, critical review of existing background information is essential. Surveys of the general biomedical literature suggest that over 50% of published papers have obvious statistical errors, and in some cases the conclusions are not supported by the published data (9–12). Such papers should not be accepted at face value. They may also cause another researcher to select inappropriate strategies and designs.

Time pressures on experimenters

There must be adequate time allowed for the completion of a project. Animals from an outside supplier need to be acclimatised for about two weeks (13), to enable them to adapt to the new environment, diet and microflora, otherwise they may be physiologically and immunologically abnormal. Techniques should have been optimised before the project is started. If staff need to learn manual skills, such as dosing procedures or surgical techniques, during the course of the experiment, this could introduce an unacceptable level of uncontrolled variation which could obscure treatment effects. There may be little apparent incentive for the researcher to reduce the number of animals, and there is sometimes a "comfort factor" in using large numbers, as it is hoped that this may obviate the need to repeat the experiment.

Pilot studies, using a few animals with the objective of determining whether a previously described model can be replicated in a new environment, are important for overall reduction. Such studies can reveal any hidden problems with dose rates or logistics, they may reveal scope for refinement, such as the choice of a more-humane endpoint, and they can provide data which can be used for estimating required sample sizes. Animals are more likely to be used unnecessarily by launching straight into a full-scale experiment, yet this is common practice.

Teamwork

Animal research is multidisciplinary, requiring expert input from research scientists, animal handlers, those concerned with animal welfare, biometricians, and possibly specialists in informatics. Procedures should be developed to allow these people to communicate effectively with each other (14). This will require written protocols and meetings to ensure that the project is feasible and can be done efficiently to the highest scientific standards.

A statistical approach to strategy

Muller *et al.* (15) provide a good basis for considering statistical aspects of research strategy. They recommend "top down planning", which involves five steps: a) specification of the experimental questions of interest; b) specification of testable hypotheses implied by these questions; c) specification of "target analyses", i.e. the statistical computations which will be necessary to estimate the presence, and size, of any treatment effects arising from the hypotheses to be tested; d) determination of the data sets which will be needed to enable such computations; and e) specification of the information which must be collected to provide the raw data. They also make a distinction between "confirmatory" experiments/analyses, which are designed to test a particular hypothesis that has been explicitly stated at the design stage, and "exploratory" ones, which explore or "mine" the data for unexpected or interesting information (15).

Exploratory data analysis should be encouraged, provided it is recognised that it gives biased estimates of statistical significance. For example, selecting the highest and the lowest mean values and performing a t test to see whether they differ significantly is unlikely to give the correct results if the experiment involves several treatment groups, unless an appropriate correction is made. This is because the t test and associated probability levels are only designed for analysing experiments which involve a comparison of the means of two groups defined before an experiment is undertaken. However, the next experiment could be designed specifically to compare two such mean values.

In some cases, the same experiment can be both confirmatory and exploratory. This is known as the "leapfrog" approach, in which each study is used to investigate a specific hypothesis, and also to generate new hypotheses for further study (15). As already noted, pilot studies can be used to gather

preliminary data which can then be used in the design of more-definitive studies.

Some complex sets of data involve measuring several different parameters for each individual. For example, haematology studies will provide data on red and white blood cell counts, packed cell volume, platelets, reticulocytes, etc. Each parameter could be analysed separately, or a multivariate analysis could be used to analyse the whole data set in a single analysis, taking account of any correlations between various parameters (16). With such complex data, it is often difficult to specify the hypotheses to be studied, and many of the multivariate statistical methods, such as principal components analysis, are essentially exploratory (17). Exploratory methods can also be used with the analysis of variance (ANOVA [18]).

Whether or not the approach suggested by Muller *et al.* (15) would suit all projects is open to debate. However, all research projects should be subject to strategic review at which the Three Rs are considered both before the experimental work commences, and periodically throughout the project. This could lead to savings of resources as well as to a reduction in animal use.

Experimental Design

Research projects often involve many separate experiments performed either sequentially or in parallel. Some of these can be uncontrolled and qualitative with a clear objective which may or may not be achieved. For example, a project may involve the production of a transgenic mouse strain which is either successful or unsuccessful. Success may depend on many of the factors discussed previously in the section on research strategy.

Controlled studies

Many experiments involve comparative studies ("controlled" experiments), in which two or more groups are compared which, as far as possible, only differ with respect to one or more treatments. These studies are capable of detecting quite subtle treatment effects, and are widely used in safety evaluation where the aim is to define the conditions under which exposure to a chemical has little or no effect. However, they need to be carefully designed if animals are not to be used unnecessarily. Many controlled studies could be improved by quite modest changes in experimental design or in the statistical analysis of the results (19).

Even the definition of what constitutes an "experiment" is not always clear. It is not unknown for a research worker to build up experimental data from a control group and some treated groups, with animals being added on an *ad hoc* basis without any a priori indication of what the eventual set of data will look like. The problem with this approach is that it assumes that the environment, experimental animals and measurement conditions remain constant. If this is not the case, any treatment comparisons will be confounded by these environmental variables. Thus, all experiments should be fully planned before any data are collected, and the experimental plan should define the treatment groups, species, strain, sex, age and number of animals, manner of randomisation, experimental design, time-scale, data to be collected, and proposed method of statistical analysis. Only in exceptional circumstances should the plan be modified once the experiment has started. Thus, it might be acceptable to eliminate a top dose group if the test chemical is unexpectedly toxic, although this could alter the power of the experiment. However, it would not usually be acceptable to add another dose group once an experiment is under way, because in such circumstances proper randomisation is impossible and there is no assurance that environmental variables will not have changed.

Some experiments could be improved by using more treatment combinations. Mead (20) suggested that most controlled experiments should involve 10–50 treatment combinations (usually in a factorial design) if resources are to be used efficiently, although this must be done with the appropriate statistical analysis. In one survey, only 10% of papers published in two toxicological journals had ten or more treatment combinations (21). Thus, if Mead is correct, there is scope for obtaining more information at little cost in terms of animal welfare or financial considerations.

Some researchers place great value on historical data, but this must be used with great care in view of the many factors that can influence a biological response (22). As noted earlier, factorial experimental designs (Figure 1), in which two or more factors (for example,

treatments, time, sex, strain, age, or diet) are varied simultaneously, usually make more efficient use of resources (including experimental animals) than do designs involving only a single factor (23–25). Where there are many factors which can influence a response, it is even possible to use fractional factorial and so-called "confounded" designs to explore their importance (26), though such designs are rarely used in biomedical research.

It is important to identify the "experimental unit", i.e. the unit which can be assigned at random and independently to a treatment. This may, for example, be an individual animal, a cage of animals, an animal for a specified time-period, or a part of an animal. The appropriate statistical analysis cannot be carried out unless the experimental unit is correctly identified. For example, if an experiment is designed with all the control animals in one cage and all the treated animals in another, the animal cannot be the experimental unit because animals in the same cage may have a common environment, so they are not independent of one another. A statistical analysis based on the assumption that the animal was the experimental unit could show whether or not the means of the two groups differed, but it would not be clear whether this was caused by the treatment or by environmental differences between the cages, possibly as a result of fighting in one, but not in the other, cage. With such a design, the cage is really the experimental unit, and no valid statistical analysis can be conducted because there are only two units in the experiment.

Reducing variability

The importance of uniformity of the experimental material cannot be over-emphasised, as it determines the extent to which treatment groups will be similar at the start of the experiment. Research workers often go

Figure 1: Example of a hypothetical 2 × 2 × 2 factorial experiment

A factorial arrangement of treatments				
	Control		Treated	
	Diet A	Diet B	Diet A	Diet B
Male	n	n	n	n
Female	n	n	n	n

The main interest might be in comparing the control and treated groups for a particular quantitative parameter. In this hypothetical example, it has been decided to incorporate two different diets and both sexes; "n" is the number of animals in each subgroup. Unequal group numbers can be accommodated with modern statistical analysis packages, although a completely blank cell would create problems. Note that the comparison between the treated and control group would probably require 12–22 animals, even if a single sex and diet were to be used (as estimated by Mead's "resource equation" method; see text). The full factorial design can similarly be done with n = 2 or 3, giving a total of 16–24 animals. However, the factorial design normally provides more information because it shows the extent to which any treatment difference depends on the sex and diet of the animals.

Almost any factor which it is thought could influence the response can be used. For example, instead of two diets, it would have been possible to use two time-points or strains, or another type of treatment, etc. Similarly, the factorial arrangement of treatments can be carried out as a randomised block design by, for example, letting n = 1 and repeating the mini-experiment, say, three times as three blocks.

to great lengths to obtain animals of uniform weight and age (often leading to unnecessary wastage, since such heterogeneity can often be accommodated by blocking), so that after random assignment to treatment groups the mean weights and ages are very similar. However, for some reason, the genetic heterogeneity found in outbred stocks is often considered to be advantageous (27), even though it means that treated and control groups are more likely to differ genetically at the start of the experiment than if more-homogeneous animals had been used. Larger numbers of animals must then be used to compensate for these differences. Where genetic variation in response is considered to be important, it should be incorporated into the experimental design by using several different strains, stocks or breeds with a factorial experimental design (2). This can be done without increasing the overall total number of animals. Differences between stocks are usually much greater than differences between individuals within a single stock, and therefore this will often result in a much wider range of susceptibility phenotypes than the use of a single heterogeneous stock (27).

Those factors which could influence the outcome of the experiment, such as the genotype, sex and age of the animals, and sources of uncontrolled variation, such as measurement error or time and space variables, need to be identified. For example, many behavioural and physiological parameters can vary with the time of day due to circadian rhythms (28, 29). Even barometric pressure can affect animal behaviour (30). Many of these factors cannot be standardised, but often can be controlled by using randomised block designs (26). A randomised complete block design is one where the experimental unit (for example, the animals) has been placed into smaller, more-homogeneous subgroups, which can be kept together throughout the experiment to minimise variation due to non-homogeneous material and time and space variables (Figure 2). Such designs often lead to substantial increases in precision at no extra cost. They are widely used in agricultural research, but not by researchers using laboratory animals. Often the experimenters do not know what is causing the variability in their studies. For example, it may be animal-to-animal, day-to-day, sample-to-sample or measurement-to-measurement variation. There are special designs ("nested designs") which can be used to identify the sources of variation. Action can then be taken to control the variability, rather than simply increasing the number of animals.

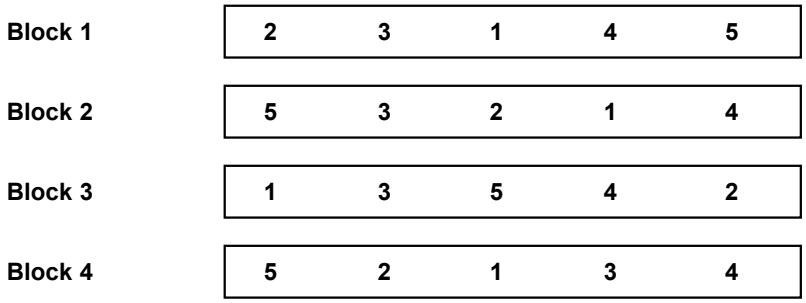
Size of the experiment

Methods of determining an appropriate size for an experiment are not widely understood. This is not surprising, as this is an area of statistics which is complex and has not yet been solved satisfactorily for all situations by mathematical statisticians. One approach is to use power analysis (31). In the past this has been difficult, as the calculations are complex for experiments with more than two treatment groups. The availability of computer programs for estimating sample sizes, such as nQuery Advisor (32), has partially solved this problem. A power analysis requires: a) an estimate of the effect size likely to be of scientific interest; b) an estimate of the standard deviation (SD); c) specification of the desired power (i.e. the chance of detecting a specified treatment effect); and d) specification of the significance level to be used.

The comparison of two laboratory animal diets, a standard diet and a new formulation designed to reduce obesity, with the body weight of male mice after they have been on the diet for six months being the dependent variable, can be used as an example. From previous work, it is known that the mean body weight ± SD of this strain of mice at six months is 44 ± 3.8g. Suppose it was specified that the result would be of interest if the mice on the new formulation weighed 15% less (i.e. 37.4g) than those on the standard formulation, and that Student's t test with a significance level of 5% and a power of 80% were to be used. Necessarily, this specification is somewhat arbitrary. By using these figures, nQuery Advisor indicates that the experiment can be done by using five mice per group. However, 18 mice per group would be required if the effect size was a reduction in weight of 7.5% compared with the mice on the standard diet.

The results of a power analysis are highly dependent on the specifications, particularly if a small effect is to be detected. Specification of an effect size of potential interest might not be too difficult with a simple experiment such as the one outlined above, but would, for

Figure 2: Diagram of a randomised block experimental design with four blocks and five treatments



The numbers in each block are codes for the treatment given to each animal. The purpose of such a design is to reduce heterogeneity associated with time, space and some physical variation. The five animals in Block 1 would be chosen to be as similar as possible with respect to age, weight, genotype and any other variable which might influence the outcome. They would then be assigned at random to one of the five treatments. The animals would be housed in close proximity, and possibly even in the same cage, if the different treatments can be given in such a situation. When measurements are to be made on the animals, those in Block 1 would be treated as a group, with all measurements being made within a short period by the same person. Animals in the other blocks would likewise be selected to be as uniform as possible, but might differ in body weight or age, etc. from those in Block 1. They could be measured at a different time, if necessary by a different person. Thus, within each block any comparison among treatments would be made on animals which are in all possible ways as similar as possible. The unwanted variation shows up as differences between blocks. This is then removed mathematically in the statistical analysis.

The randomised block design provides a means to break down an experiment into smaller parts, which can be handled more conveniently. In most cases, it will increase the precision at no extra cost, apart from the need for a slightly more-complex statistical analysis.

example, be difficult for a factorial experimental design with several treatment combinations and dependent variables. An estimate of the SD could be obtained from a previous study, from the literature, or from a pilot study. However, all these sources of information are subject to error. The significance level is usually set, somewhat arbitrarily, at 0.05, and the power is likewise often set at 80–90%. Thus, although doing a power analysis is a useful exercise in showing the potential capability of various proposed experiments to detect an effect of biological interest, it does not always give a definitive answer of the most appropriate size of each experiment. It is strongly recommended, therefore, that sample sizes are continually reviewed as experimental results become available.

An advantage of a power analysis is that it can be used to explore the implications of negative results, i.e. those in which there is no significant difference between treatment means. Such negative results might be of biological interest, particularly in safety testing, if they are real, but are of little interest if the lack of statistical significance was because the experiment was too small to detect a treatment difference of potential interest. Thus, a power analysis can be used to find out the probability that the experiment would have been able to detect a specified treatment effect if it was really there. For example, in the diet experiment discussed previously, suppose that an experiment had been undertaken to compare the two diets with ten mice per group, and that the mean

body weights at the end of the experiment had been virtually identical at 44g with an SD averaged across groups of ± 4g. The statistical analysis would indicate that there was no significant difference in body weight between the animals on the two diets, but would not indicate whether the experiment could have detected a difference of biological interest. A power analysis could be used to determine the probability that this experiment would have detected a 15% change in body weight given that the mean weight of the controls was 44 ± 4g, and there were ten mice per group. This is easily done by using nQuery Advisor, which indicates that with such an experiment there would have been a 93% chance of detecting such a treatment effect if it existed. Thus, if the diet was really capable of reducing body weight by 15%, there would have been a good chance of detecting it.

The alternative "resource equation" method for determining the size of an experiment (20), is a rule-of-thumb approach based on the observation that, for experiments involving quantitative variables, diminishing returns of information are found if the size of the experiment is increased so that there are substantially more than about 20 degrees of freedom for the error term, "E". However, good returns are found from using more animals if there are less than about 10 degrees of freedom for error. Thus, the optimum size of an experiment usually has between 10 and 20 error degrees of freedom. As an example, the earlier experiment involving a new formulation of a mouse diet compares the mean body weights of the mice by using an unpaired t test which has E = n – 2 error degrees of freedom, where n is the total number of animals. Thus, for E = 10–20, the experiment should use a total of 12–22 mice, or 6–11 mice per group, which is in broad agreement with estimates from the power analysis.

The resource equation method is easy to use with quite complex experimental designs. For completely randomised designs (i.e. not randomised block designs), E is the total number of animals minus the total number of treatment combinations; so, assuming a completely randomised design in Figure 1, if n = 4, there will be a total of 32 animals and eight treatment combinations, so E is 32 – 8 = 24. If n is reduced to 3, E will be 24 – 8 = 16. With randomised block designs, the number of blocks less one also needs to be subtracted. For example, in Figure 2 there is a total of 20 animals, with five treatments and four blocks; in this case, E = 20 – 5 – 4 + 1 = 12. Note that for experiments with only two treatment groups, group size might appropriately be between six and 11 to give E = 10–20, whereas for larger experiments such as the factorial experiment shown in Figure 1, which has eight treatment combinations, a group size of only three would be required to give E = 16. Thus, group size can be reduced and more information can usually be obtained if there are several treatment groups. Note that the use of blocking might appear to be counter-productive as it reduces E. However, the reduction in the error variance when using a randomised block design usually more than compensates for this (19).

As the resource equation method does not specify statistical power, effect size of interest, SD or significance level, it will not be known in advance how effective the experiment will be in detecting a particular treatment effect (21). It will be known, however, that little would be gained from using substantially more animals than the numbers required to give about 20 degrees of freedom for error. However, for large and complex experiments, the upper limit of 20 degrees of freedom may be so restrictive that it is impossible to have a balanced experiment with the same numbers of animals in each group, which is also desirable. Therefore, the limits of E = 10–20 should not be applied too rigidly. Also, for some *in vitro* tests where, for example, the experimental unit might be a tissue culture dish, including more dishes could be inexpensive. In such circumstances, it might be economical to allow E to be much higher than would generally be acceptable if animals were being used. Having done some experiments applying the resource equation method, it might be desirable to explore their power characteristics by using a power analysis.

Designing experiments with excessive numbers of animals resulting in unnecessarily high precision should also be avoided. The existence of very low probability (p) values (for example, $p < 0.001$) indicates that the experiment may have been unnecessarily large. Hendriksen *et al.* (33) found that assays of adsorbed diphtheria and tetanus vaccine could usually be undertaken with

half the number of animals presently required, yet still be within the limits of confidence stipulated by the European Pharmacopoeia and the World Health Organization (WHO). They also suggested that there should be some flexibility in national and international requirements to allow for individual circumstances.

Sequential experimentation

Sequential designs (34) in which the outcome of an experimental (and control) treatment is observed with small numbers of animals (in a "mini-experiment"), followed either by reaching a conclusion about the effect of the treatment or by taking a decision to treat another sample of animals, could be more widely used. Indeed, there may be scope for the more-widespread use of such designs in experimental surgery (35). Sequential designs often use substantially fewer animals than those involving fixed numbers (36), but they are only applicable to relatively simple experiments where the results are quickly available. The main limitation is that the results of each mini-experiment must be available before the next one is undertaken, so that the appropriate decision can be taken.

Sequential designs for evaluating the LD50 or ED50 (50% effective dose) of a compound have been known for many years, but are sometimes difficult to apply. Two improved sequential approaches have been described recently. The Fixed Dose Procedure (37) and the Acute Toxic Class method (38) have been evaluated in some detail and both use fewer animals than conventional LD50 tests; in addition, they can incorporate observations of toxic symptoms rather than using death as the endpoint. Wherever possible, an ED50 using a more-humane endpoint than death, such as a change in behaviour or a specified reduction in body temperature (39), should be used, even though it may not strictly indicate acute lethal potential.

A special case of sequential experimentation is to use a Bayesian approach, where the researcher's prior beliefs about the outcome of an experimental treatment are updated and modified by the availability of sequential sets of new experimental data. Unfortunately, Bayesian statistical methods are not discussed in most elementary statistical text books, and even introductory texts (40) have a highly mathematical approach which is inaccessible to most experimentalists. Thus, this approach would normally require the active involvement of a professional statistician.

In conclusion, there seems to be considerable scope for improving the design of individual experiments to reduce the number of animals needed for a given research output. A person with expertise in experimental design should be involved in planning experiments, with this involvement being formally recognised.

Statistical Analysis

The aim of statistical analysis is to extract all useful information from the data. The method of analysis will be closely linked to the experimental design and to the type of data to be produced. Researchers should normally have a clear idea of how they intend to analyse the results at the experimental design stage.

Common problems

A common statistical mistake seems to be the use of Student's t test to analyse experiments which have more than two treatment groups (41). In such circumstances, the t test may lack statistical power, so that real treatment effects can be missed. It can also lead to false positive results if many different comparisons are made, and it is not easy to test for potentiation or interaction, such as a different response to a drug treatment in males and females, in a factorial experimental design. Other common mistakes include failure to take account of variation among heterogeneous experimental groups, and failure to present any statistical analysis even though numerical data are generated (21).

Solutions

In practice, most measurement data from controlled experiments can be analysed by using the ANOVA, a highly versatile technique which can be used to analyse quite complex data sets. The method requires the assumption that the residuals (i.e. the deviations from the group means) are independently and normally distributed and are the same in each group, although a scale transformation can be used to achieve these conditions. Most statistical packages now

provide diagnostic methods for studying these residuals, and scale transformation is easy, should it be necessary. Although factorial designs are commonly used, they are not always correctly analysed in terms of the marginal means of each factor, and the interactions between factors. In view of the importance and value of factorial designs, more training in their use and analysis might be appropriate.

Many other statistical methods are potentially useful and could be more widely used, including various tests for comparing proportions, tests for trends and correlations, and multivariate methods such as principal components analysis for analysing data where there are several dependent variables (16).

In conclusion, there appears to be considerable scope for better statistical analysis of experiments as a means of extracting more useful information which, in the long run, should reduce animal use.

Interpretation and Communication

The results of each experiment need to be interpreted, and in many cases the design of the next experiment depends upon this interpretation, which is sometimes flawed. A common error is to base the interpretation on statistical significance (usually a p value) rather than on the magnitude of the treatment effect. A treatment effect can be statistically significant but of little biological interest and, conversely, a biologically interesting effect might not be statistically significant because the experiment was poorly designed and unable to detect it. As an example, a paper submitted for publication (and rejected in its present form) claimed that in a genetically heterogeneous population exposed to a carcinogen, the levels of DNA adducts (a measure of DNA damage) were significantly associated with genotype at a polymorphic locus. However, the paper was written in such a way that it was impossible to determine whether genotype was numerically important or was just one of many factors that affected the adduct levels. Thus, the question that was answered, as a result of undue emphasis on p values, was "does genotype affect adduct levels?", but the question of real interest, which could have been answered equally well with the available data, was "to what extent does genotype influence adduct levels?"

The results of any experiment should be clearly presented by using suitable tables and graphs. The presentation of tables, in particular, could often be improved. It is almost universal practice to quote a mean with an SD or SE based on the animals within that particular group, even though a pooled SD across groups would provide a better estimate of the population SD (41). If pooled SDs were more widely used, the means could be presented much more clearly without each one having ± SD appended to it. Journal editors and referees could suggest such modifications and also play an important part in improving the statistical quality and presentation of published papers. Papers should generally be sent to a specialist statistical referee, or a biologist with a good understanding of statistics, if they contain any numerical data. If editors have difficulty finding statisticians prepared to referee papers, they should consider offering a fee. Guidelines on statistical analysis have been published in a few journals (15), although these are difficult to develop in view of the wide range of methods that might be used. However, all journals publishing papers on studies which involved the use of experimental animals should include a statement requiring authors to adhere to strict humane standards, pointing out that this implies the use of the minimum number of animals needed to achieve the scientific objectives of the study (42).

In conclusion, there is scope for improving the interpretation and presentation of the results of individual experiments. This would improve the communication of research results, and could also lead to a reduction in animal use.

Legislation and Internal Review

Legislation and internal review procedures employing inspectors and/or ethics or animal experimentation committees have been developed, in part, as a response to the demand for the use of humane techniques. However, the law and requirements for review procedures vary between countries, and between institutes or companies within a country.

Laws relating to the use of animals in the European Union (EU) Member States have been enacted in response to *Directive 86/609/EEC* (43). Article 7(3) states that: "In a choice between experiments, *those which use the minimum number of animals*, involve animals with the lowest degree of neurophysiological sensitivity, cause the least pain, suffering, distress or lasting harm *and which are most likely to provide satisfactory results* shall be selected" (italics added for emphasis). Thus, in the EU, there is a clear requirement that reduction should be considered as an integral part of the review process, although countries differ in the exact wording of their specific national legislation.

Similar legislation has either not been introduced, or has not been enforced, in many countries worldwide. This has resulted in many countries lacking a legal requirement to use the minimum number of animals. Reviews of the development of alternatives in relation to the legislation in force in various countries indicate that approaches vary considerably (44–48). The establishment of some kind of ethics review committee seems to be common, although the composition, remit and effectiveness of such committees varies. Whatever their composition, such committees usually assess the quality of the proposed project, and often suggest improvements which could result in a reduction in animal use.

Adherence to regulations and guidelines not primarily designed to promote animal welfare, such as those associated with Good Laboratory Practice and international standards such as ISO 9001, might also lead to a reduction in the use of animals, because they ensure that procedures are carried out consistently, to predefined standards which are less likely to be erroneous or inappropriate (49). However, there is a danger that such regulations could prove inflexible, and could, in some cases, result in animals being used to satisfy bureaucratic, rather than scientifically justifiable, objectives.

The success of attempts to reduce the use of animals as a combined result of replacement and reduction initiatives should be monitored, although this presents problems. Few countries collect accurate statistics on laboratory animal use, and in no case can the use of laboratory animals be related to research output. Even within the EU, where *Directive 86/609/EEC* provides a common legislative framework, there is considerable variation between Member States in the collection of statistics on animal use. Some practical proposals have been made to standardise the statistics, including some modifications to the Directive to remove ambiguities, the use of a standard set of tables by all Member States, the use of legal measures to enforce adequate data collection, and the development of some methods of quality control to ensure the accuracy of the data (50). These suggestions need to be implemented.

Reduction in animal use should also be related to research output. Strong emphasis on total numbers and the setting of arbitrary targets should be avoided, as they can be counter-productive in terms of animal welfare. For example, such emphasis could result in excessive re-use of animals, at the expense of their welfare. Pharmaceutical research and chemical production are increasingly conducted by multinational companies, and research can be relocated to another country if the laws regulating animal use in a particular country were to make it too difficult. It would be counter-productive, in terms of animal welfare, if there was too much migration of research and testing to countries with less stringent animal welfare regulations.

International harmonisation (via the International Conferences on Harmonisation) of standards governing the toxicity testing of pharmaceuticals appears to have resulted in nearly a 50% reduction in the number of animals which are used to test some pharmaceuticals (51). If this reduction is realised across the board, it will also represent a considerable financial saving. Such harmonisation should continue and be extended to the testing of other chemical and biological compounds. Guidelines such as those produced by the OECD should be updated periodically, with emphasis being placed on the Three Rs, if possible by using an external ethical review panel. Where improved methods are introduced, their adoption should be promoted, and older guidelines should be deleted after a suitable period of time. Where full harmonisation cannot be achieved, mutual recognition of data by national control or regulatory authorities should be adopted, to avoid the duplication of tests.

In conclusion, there appears to be scope for greater international harmonisation of animal welfare legislation, methodology for internal review, and for the continued harmonisation of testing procedures for pharmaceutical, chemical and biological compounds, with specific emphasis being placed on the implementation of the Three Rs.

Education and Information

The provision of suitable education for researchers, animal house staff, members of ethics committees, and others involved in animal research should result in the use of fewer animals for a given research output (52). However, there is considerable variation between countries, and between institutes within each country, in the provision of such education. In some developing countries, there is no training available for animal technicians or veterinary professionals, and there is no requirement or possibility for researchers to attend courses on humane techniques in their own country. A relatively small additional effort, and financial support, from developed countries working through international organisations such as the International Council on Laboratory Animal Science and the WHO could lead to considerable improvements in laboratory animal welfare standards and the quality of animal experimentation in these countries, in conjunction with a reduction in animal use.

Within developed countries, such as those in the EU, there is usually a well-developed career structure for those working with animals, with a legal requirement for the provision of veterinary staff in laboratory animal facilities. Researchers are also required to undergo training in the handling and use of experimental animals. Several working parties established by the Federation of European Laboratory Animal Science Associations (FELASA) have considered the training of people using experimental animals. Four categories of such people were recognised (53): A — those taking care of animals; B — those carrying out animal experiments; C — those responsible for directing animal experiments; and D —laboratory animal science specialists.

The working party covering categories A and C published its report in 1995, including a teaching syllabus (discussed in reference 53). The syllabus for people in category C includes some training in experimental design and statistics, which would serve to improve communication between researchers and statisticians. Although such courses have now been running for several years, it will take some time before all research scientists have been trained. Research should now be undertaken to determine how effective the training has been in improving the quality of research and in reducing animal usage, how the training can be improved, and whether refresher courses are needed.

The potential benefits of reducing the number of animals used in a given project are not always appreciated by research scientists. While the financial impact of using small numbers of additional animals appears to be negligible, research progress is often limited by the resources available. In some cases, researchers spend considerable amounts of time reading slides or making measurements on tissues taken from animals. Smaller experiments would save time and resources in addition to animals and, provided the experiments are well designed, research progress would be more rapid. Researchers might be more receptive to the concept of reduction if the wider economic benefits were realised.

Ideally, each project team should have access to statistical advice. However, communication between biologists and statisticians is often unsatisfactory, and in some (usually academic) institutes there may be a consultancy fee for statistical advice. This is a strong disincentive for the researcher to consult a statistician, with the possible result that animals are used unnecessarily, for the reasons discussed previously. Moreover, there is a need for improved communication between statisticians and researchers. If statisticians were consulted more frequently, and in some cases were included as joint authors, they could more easily acquire an understanding of the practical aspects which have to be taken into account when optimising the design of animal experiments. Funding authorities, and the institute where the work is carried out, should be required to provide sufficient resources to ensure that research is conducted to high standards, and this should include sufficient funds to allow input from a statistician in cases

where the researcher needs such advice. If the regulatory authorities administering animal welfare legislation are not satisfied that adequate statistical advice is available at a cost which the researchers are able to pay, the authority to carry out laboratory animal research on those premises might have to be withdrawn.

The possible value of computer-aided learning of experimental design and statistics needs to be evaluated. A number of programs have been developed for teaching statistics, but they are not all relevant and none has yet been used as a means of implementing the Three Rs. However, the use of such learning aids might help to cater for the heterogeneous backgrounds of researchers working with animals. Some home study, done in conjunction with formal teaching, might provide an economic way of increasing the knowledge of experimental design and statistics of researchers, and might provide useful material for refresher courses.

Conclusions and Recommendations

Reduction through the application of better research strategy, experimental design and statistical methods

1. All projects which might involve the use of experimental animals should be reviewed at regular intervals, to include consideration of how reduction, refinement and replacement are to be incorporated in the experimental matrix or strategy. The review panel should include at least one person independent of the research group undertaking the work.
2. Guidelines and/or checklists should be developed to assist this review process, and methods should be developed to monitor the success or failure of the project reviews in reducing animal use.
3. Workshops covering research strategy should be organised within companies, industry associations, academic institutes and scientific societies every 3–5 years. This topic should be supplementary to any coverage of experimental design, and should presume knowledge equivalent to FELASA's Category C syllabus for research scientists.
4. Lists of suitable reference literature and computer programs should be developed to resource such workshops.
5. To ensure that the optimum number of animals is used, a person with expertise in experimental design should be involved in planning all experiments, and this involvement should be formally recorded.
6. Guidelines should be developed for implementing Recommendation 5, which should include ways in which feedback and improvement might be incorporated into the consultation process.
7. The education and training of experimenters should include discussion of the types of experimental design and their applications. The experimenters should be actively involved by using real case studies wherever possible. Training courses should aim to bring experimenters to a level where they can communicate effectively with experts in experimental design, and should include an awareness of the range of available experimental designs and statistical analyses, and of the interpretation and presentation of results. Those involved in the review process outlined above should be informed of the level of training in experimental design achieved by each researcher.
8. Resource materials, including a syllabus, for such courses should be gathered and/or developed.
9. Journal editors should be encouraged to require authors to provide brief descriptions of the type of experimental design used, and to improve the presentation of data and their analysis in publications.

Legislative and organisational framework

10. Laws to protect laboratory animals and to encourage the development of high ethical standards in the use of animals should be enacted in those countries where no such laws currently exist.
11. There should be increased support for developing countries in establishing guidelines, legislation and educational programmes in relation to the use of experimental animals and the Three Rs,

with particular reference to the potential scientific and economic benefits from improving the quality of biomedical science. This could most effectively be implemented through organisations with suitable links with such countries, such as the WHO and the International Council for Laboratory Animal Science.

12. Where there are scientific and academic exchanges between developing and developed countries, information and discussion on the Three Rs should form part of the study and academic programmes.

13. Where laws exist to protect laboratory animals, national authorities should ensure that they are effectively implemented.

14. All institutes where experimental animals are used should be required to maintain and document internal review processes that specifically address the implementation of the Three Rs. Examples of people who could be involved in this process include animal technical and veterinary staff, those with knowledge of alternative methods, those with expertise in statistics in relation to the needs of biological projects, and people who are independent of the work of the institution.

15. Information on internal review processes, adopted either as a consequence of national legislation or voluntarily, should be collated to assist in further developing these processes.

16. All EU Member States should be required to produce annual statistics on the use of experimental animals which are accurate, comprehensive and comparable. Research should be undertaken to identify methods of measuring trends in animal use relative to scientific output.

17. International harmonisation of testing procedures should be a continuing process, and should specifically address ways in which the Three Rs can be further implemented.

Education, training and information

18. All research workers who use experimental animals should have appropriate training in, for example, the biology of the species to be used, possible microbiological hazards to humans and animals, the design and conduct of experiments, anaesthesia, analgesia, experimental techniques, replacement alternatives, ethical aspects, and analysis of scientific papers. Such training could be based on the curriculum proposed by FELASA, but with adjustment for individual circumstances.

19. In training scientists in the humane use of laboratory animals, the positive benefits in terms of improved scientific quality and output resulting from reducing the numbers of animals used (as well as from the consideration of other alternatives) should be stressed.

20. Funding authorities and organisations involved with animal research should ensure that the necessary resources are available for conducting humane research. These facilities should include access to training and education for all categories of staff, and the provision of adequate statistical advice at a reasonable cost.

21. Communication between researchers and statisticians/biometricians needs to be improved. Consideration should be given to the development of training courses or workshops on practical and theoretical aspects of animal experimentation for non-clinical statisticians/biometricians.

22. An evaluation of computer-aided learning courses, databases and information sources which could assist in implementing reduction should be undertaken, and the results made available to teachers involved in training scientists in humane techniques.

23. Education is a continuing process. Refresher courses should be offered to reinforce and update information. All courses should be taught in a flexible manner to take account of, and benefit from, the heterogeneous background of most participants.

General recommendations and possible methods for implementation

24. The recommendations and conclusions of two previous workshops, on "The

Three Rs: the Way Forward" (54; Table I) and "Guidelines for Reviewing Manuscripts on Studies Involving Live Animals" (55; Table II), which are relevant to this workshop, should be reviewed in terms of the progress made with their implementation. Where progress appears to have been unsatisfactory, ways should be sought to implement the recommendations more effectively.

25. A Standing Committee on Reduction, of 5–10 members with suitable expertise in laboratory animal science and technology, statistics, and biomedical education and information, should be established under the auspices of an appropriate governmental or charitable organisation, specifically to progress the recommendations in this and other reports.
26. A person should be appointed, initially on a short-term contract, to support the above committee and to provide human resources for implementing those recommendations where immediate intervention might be successful.

References

1. Anon. (1994). ECVAM News & Views. *ATLA* **22**, 7–11.
2. Russell, W.M.S. & Burch, R.L. (1959). *The Principles of Humane Experimental Technique*, 238 pp. London: Methuen.
3. Flecknell, P.A. (1994). Refinement of animal use: assessment and alleviation of pain and distress. *Laboratory Animals* **28**, 222–231.
4. Festing, M.F.W. (1997). Variation and its implications for the design of experiments in toxicological research. *Comparative Haematology International* **7**, 202–207.
5. Gruneberg, H. (1952). *The Genetics of the Mouse*,

Table I: Reduction alternatives: conclusions and recommendations of ECVAM workshop report 11[a]

13. In cases where a choice between species is possible, there is generally no scientific justification for using more of the smaller species than of the larger one.

14. Research strategy should be considered carefully, with a view to reducing the numbers of animals used. The example of Hendriksen *et al.* (7), in choosing strains of laboratory mice in order to minimise the numbers needed in specific biological assays, should be followed for those assays which use large numbers of animals and which are unlikely to be replaced with *in vitro* alternatives in the near future.

15. The design of regulatory testing procedures, including the sample sizes required, should be reviewed regularly, possibly as part of international harmonisation.

16. Substantial reduction in animal use could be achieved by further harmonising toxicity testing regulations, for example, with respect to group sizes, dose levels and the length of studies.

17. In view of the uncertainties inherent in "extrapolating" to humans, the need for very high precision in data provided by animal experiments should be reconsidered.

18. There is evidence that some non-regulatory animal experiments are poorly designed and incorrectly analysed. As a minimum, all research workers should have adequate training in experimental design and the proper use of statistical methods.

19. The concept of the "named statistician" as an essential part of the regulatory framework of animal experimentation should be explored.

[a] *Taken from Balls et al. (54).*

Table II: Conclusions from a workshop on guidelines for reviewing manuscripts on studies involving live animals[a]

1. All journals publishing papers which might involve animal suffering or distress should be encouraged to have a statement of journal policy with respect to the use of research animals. This should normally be published in the instructions to authors.

2. No single policy statement is appropriate for all journals.

3. A set of example statements should be developed, which could be made available to editors, in order to assist them in developing or enhancing an appropriate policy.

4. A policy statement alone is generally not sufficient to ensure that it is followed. All referees should have a copy of the guidelines for authors, so that failure to comply is more likely to be noted by them.

5. Editors could consider requiring authors to sign a declaration that they have followed the appropriate ethical procedures, or alternatively, an appropriate statement of ethical compliance should be included in the journal article.

6. The paper should contain some justification for the use of animals, stating why no alternative approach could be used.

7. In some cases, papers do not give sufficient information about the animals to enable other research workers to repeat or correctly interpret a study. A checklist of information which might be appropriately recorded in the materials and methods section of a particular paper could be helpful to journal editors. It is not suggested that all the information on the checklist would be used in every paper, but it would act as a reminder to authors, editors and referees of the need to provide sufficient information.

8. It is not possible for an *ad hoc* working party to develop and disseminate the appropriate material to journal editors. This will require effort over quite a long period, and a modest budget.

9. Dissemination should, as far as possible, be through existing "umbrella" organisations for science editors, such as the Council of Biology Editors (CBE), the European Association of Science Editors (EASE) and the International Committee of Medical Journal Editors (ICMJE).

10. It was agreed that ECVAM should be asked to consider establishing a working party on this subject, with the aim of implementing and extending the ideas set out in this document.

[a]*Taken from Festing & van Zutphen (55).*

650 pp. The Hague, The Netherlands: Nijhof.
6. Festing, M.F.W. (1991). The reduction of animal use through genetic control of laboratory animals. In *Replacement, Reduction and Refinement: Present Possibilities and Future Prospects* (ed. C.F.M. Hendriksen & H.B.W.M. Koëter), pp. 193–212. Amsterdam: Elsevier.
7. Hendriksen, C.F.M., Slob, W., van der Gun, J.W., Westendorp, J.H.L., den Bieman, M., Hesp, A. & van Zutphen, L.F.M. (1994). Immunogenicity testing of diphtheria and tetanus vaccines by using isogenic mice with possible implications for potency testing. *Laboratory Animals* **28**, 121–129.
8. Festing, M.F.W. (1990). Contemporary issues in toxicology: use of genetically heterogeneous rats and mice in toxicological research: a personal perspective. *Toxicology and Applied Pharmacology* **102**, 197–204.

9. Gore, S.M., Jones, I.G. & Rytter, E.C. (1977). Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976. *British Medical Journal* **8 January 1977**, 85–87.
10. Altman, D.G. (1982). Statistics in medical journals. *Statistics in Medicine* **1**, 59–71.
11. Altman, D.G. (1982). Misuse of statistics is unethical. In *Statistics in Practice* (ed. S.M. Gore & D.G. Altman), pp. 1–2. London: British Medical Association.
12. Altman, D.G. (1991). *Practical Statistics for Medical Research*, 611 pp. London: Chapman & Hall.
13. Beynen, A.C., Festing, M.F.W. & van Montfort, M.A.J. (1993). Design of animal experiments. In *Principles of Laboratory Animal Science* (ed. L.F.M. van Zutphen, V. Baumans & A.C. Beynen), pp. 209–240. Amsterdam: Elsevier.
14. Stephens, U.K. & Moore, M.T. (1997). Harmonisation of the research effort: a communication model in the USA. In *Harmonisation of Laboratory Animal Husbandry: Proceedings of the Sixth FELASA Symposium, 19–21 June 1996, Basel, Switzerland* (ed. P.N. O'Donoghue), pp. 170–172. London: The Royal Society of Medicine.
15. Muller, K.E., Barton, C.N. & Benignus, V.A. (1984). Recommendations for appropriate statistical practice in toxicological experiments. *Neurotoxicology* **5**, 113–126.
16. Festing, M.F.W., Hawkey, C.M., Hart, M.G., Turton, J.A., Gwynne, J. & Hicks, R.M. (1984). Principal components analysis of haematological data from F344 rats with bladder cancer fed the retinoid *N*-(ethyl)-all-*trans*-retinamide. *Food and Chemical Toxicology* **22**, 559–572.
17. Maxwell, A.E. (1977). *Multivariate Analysis in Behavioural Research*, 164 pp. London: Chapman & Hall.
18. Hoaglin, D.C., Mosteller, F. & Tukey, J.W. (1991). *Fundamentals of Exploratory Analysis of Variance*, 430 pp. New York: John Wiley.
19. Festing, M.F.W. (1992). The scope for improving the design of laboratory animal experiments. *Laboratory Animals* **26**, 256–267.
20. Mead, R. (1988). *The Design of Experiments*, 620 pp. Cambridge: Cambridge University Press.
21. Festing, M.F.W. (1996). Are animal experiments in toxicological research the "right" size? In *Statistics in Toxicology* (ed. B.J.T. Morgan), pp. 3–11. Oxford: Clarendon Press.
22. Roe, F.J.C. (1994). Historical histopathological control data for laboratory rodents: valuable treasure or worthless trash? *Laboratory Animals* **28**, 148–154.
23. Fisher, R.A. (1960). *The Design of Experiments*, 248 pp. New York: Hafner Publishing Company.
24. Maxwell, S.E. & Delaney, H.D. (1989). *Designing Experiments and Analyzing Data*, 902 pp. Belmont, CA, USA: Wadsworth Publishing Company.
25. Cox, D.R. (1958). *Planning Experiments*, 208 pp. New York: John Wiley.
26. Cochran, W.G. & Cox, G.M. (1957). *Experimental Designs*, 611 pp. New York: John Wiley.
27. Festing, M.F.W. (1995). Use of a multi-strain assay could improve the NTP carcinogenesis bioassay program. *Environmental Health Perspectives* **103**, 44–52.
28. Fox, R.R., Laird, C.W. & Kirshenbaum, J. (1974). Effect of strain, sex, and circadian rhythm on rabbit serum bilirubin and iron levels. *Proceedings of the Society of Experimental Biology and Medicine* **145**, 421–427.
29. Schwartz, W.J. & Zimmerman, P. (1990). Circadian timekeeping in Balb/c and C57BL/6 inbred mouse strains. *Journal of Neuroscience* **11**, 3685–3694.
30. Sprott, R.L. (1967). Barometric pressure fluctuations: effect on the activity of laboratory mice. *Science* **157**, 1206–1207.
31. Muller, K.E. & Benignus, V.A. (1992). Increasing scientific power with statistical power. *Neurotoxicology and Teratology* **14**, 211–219.
32. Elashoff, J.D. (1995). *nQuery Advisor User's Guide*, 202 pp. Boston, MA, USA: Statistical Solutions.
33. Hendriksen, C.F.M., van der Gun, J.W., Marsman, F.R. & Kreeftenberg, J.G. (1987). The effects of reductions in the numbers of animals used for the potency assay of the diphtheria and tetanus components of adsorbed vaccines by methods of the European Pharmacopoeia. *Journal of Biological Standardisation* **15**, 287–297.
34. Armitage, P. (1975). *Sequential Medical Trials* 2nd edn, 191 pp. Oxford: Blackwell Scientific Publications.
35. Festing, M.F.W. (1975). Experimental design. In *An Introduction to Experimental Surgery* (ed. J. De Boer, J. Archibald & H.G. Downie), pp. 5–45. Amsterdam: Excerpta Medica.
36. Chamove, A.S. (1996). Reducing animal numbers: sequential sampling. *AWIC Newsletter* **7**, 3–6.
37. Whitehead, A. & Curnow, R.N. (1992). Statistical evaluation of the fixed-dose procedure. *Food and Chemical Toxicology* **30**, 313–324.
38. Schlede, E., Mischke, U., Roll, R. & Kayser, D. (1992). A national validation study of the acute-toxic-class method: an alternative to the LD50 test. *Archives of Toxicology* **66**, 455–470.
39. Wong, J.P., Saravolac, E.G., Clement, J.G. & Nagata, L.P. (1997). Development of a murine hypothermia model for study of respiratory tract influenza virus infection. *Laboratory Animal Science* **47**, 143–147.
40. Lee, P.M. (1989). *Bayesian Statistics: An Introduction*, 344 pp. London: Arnold.
41. Festing, M.F.W. (1994). Reduction of animal use: experimental design and quality of experiments. *Laboratory Animals* **28**, 212–221.
42. Boisvert, D.P.J. (1997). Editorial policies and animal welfare. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen & M. Balls), pp. 399–404. Amsterdam: Elsevier.
43. Anon. (1986). Council Directive of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. *Official Journal of the European Communities* **L358**, 1–29.
44. Strandberg, J.D. (1997). National/regional developments in alternatives and animal use: an overview. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen & M. Balls), pp. 125–126. Amsterdam: Elsevier.
45. de Greeve, P. & de Leeuw, W. (1997). Developments in alternatives and animal use in Europe. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen & M. Balls) pp. 127–136. Amsterdam: Elsevier.
46. Johnson, N.E. (1997). Developments in alterna-

tives and animal use in Australia/New Zealand. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen, & M. Balls), pp. 137–143. Amsterdam: Elsevier.
47. Kuroda, Y. (1997). Developments in alternatives and animal use in Japan. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen & M. Balls), pp. 145–150. Amsterdam: Elsevier.
48. Stephens, M.L. (1997). Developments in alternatives and animal use in North America. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen & M. Balls), pp. 151–154. Amsterdam: Elsevier.
49. de Vrey, P. (1997). The use of quality assurance systems in maintaining control over animal experiments. In *Harmonisation of Laboratory Animal Husbandry: Proceedings of the Sixth FELASA Symposium, 19–21 June 1996, Basel, Switzerland* (ed. P.N. O'Donoghue), pp. 79–83. London: The Royal Society of Medicine.
50. Rusche, B., Sauer, U.G. & Kolar, R. (1997). *Evaluation of the Statistical Information Concerning the Number of Animals Used for Experimental or Other Scientific Purposes in the EU Member States According to Directive 86/609/EEC*, 46 pp. Neubiberg, Germany: Akademie für Tierschutz.
51. Lumley, C.E. & van Cauteren, H. (1997). Harmonisation of international toxicity testing guidelines for pharmaceuticals: contribution to refinement and reduction in animal use. *EBRA Bulletin* **November 1997**, 4–9.
52. Festing, M.F.W. (1981). The "defined" animal and the reduction of animal use. In *Animals in Research: New Perspectives in Animal Experimentation* (ed. D. Sperlinger), pp. 285–306. Chichester: John Wiley and Sons.
53. van Zutphen, L.F.M. (1997). Education and training in laboratory animal science. In *Harmonisation of Laboratory Animal Husbandry: Proceedings of the Sixth FELASA Symposium, 19–21 June 1996, Basel, Switzerland* (ed. P.N. O'Donoghue), pp. 160–166. London: The Royal Society of Medicine.
54. Balls, M., Goldberg, A.M., Fentem, J.H., Broadhead, C.L., Burch, R.L., Festing, M.F.W., Frazier, J.M., Hendriksen, C.F.M., Jennings, M., van der Kamp, M.D.O., Morton, D.B., Rowan, A.N., Russell, C., Russell, W.M.S., Spielmann, H., Stephens, M.L., Stokes, W., Straughan, D.W., Yager, J.D., Zurlo, J. & van Zutphen, L.F.M. (1995). The Three Rs: the way forward. The report and recommendations of ECVAM workshop 11. *ATLA* **23**, 838–866.
55. Festing, M.F.W. & van Zutphen, L.F.M. (1997). Guidelines for reviewing manuscripts on studies involving live animals. Synopsis of the workshop. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen & M. Balls), pp. 405–410. Amsterdam: Elsevier.