

Control of Variability

B. R. Howard

Abstract

Genetic variability is an important component of the phenotypic variation within populations; however, there are often many other contributing factors, which receive little attention. Randomization techniques or grouping factors known to contribute to variability can do much to isolate, or at least accommodate, this variability. The European Centre for the Validation of Alternative Methods addressed some of these approaches in 1998, and others are considered in this article. However, laboratory animals are living beings that respond to scientific procedures, or indeed to subtle variations in husbandry conditions, which our senses do not equip us to comprehend readily. These variations may either bias our experimental results or introduce sufficient background noise to mask differences arising from the scientific procedures. In planning an experiment, it is important to devote adequate time considering such factors and developing appropriate strategies to handle them.

Key Words: consistency; experimental design; experimental procedures; handling; husbandry; measurements; variability

Background

Life forms characteristically adapt to their environment. The success with which this is achieved varies from one individual to another, and hence that adaptability is a challenge to precise experimental design. It has long been recognized that biological observations are superimposed on a background of variability that is substantially greater than that of the traditional chemical and physical disciplines. In part, this variability is a consequence of the fact that life forms are chemically (metabolically) very complex; however, it is also a fact that animals are inherently less stable, due to their active responses to environmental influences. The greater the population variability, the greater the number of animals the experimenter must use to detect a given biological change.

The European Centre for the Validation of Alternative

Methods addressed some approaches to variability in a 1998 workshop (ECVAM 1998). This article explores the basis of that variability and examines ways to address it. It is also important to recognize that because animals are sentient beings, we must treat them with respect and endeavor to comply with the principles of reduction, refinement, and replacement (the 3Rs)—using animals only when necessary and minimizing the numbers required and the impact of the investigation on them.

There are three sources of variability in experiments using live animals:

1. Variability introduced by the experimenter. Examples include (a) the way in which known variables are incorporated into the experimental design, (b) how the animal is prepared for the investigation, (c) the conditions under which variables are measured, and (d) the precision with which measurements or techniques are carried out.
2. Inherent variation between animals. Examples are due to their (a) genetic constitution, (b) sex, (c) age, and (d) body weight. Sometimes this variability is relatively easy to recognize, in which case it can be controlled by choosing a uniform population of individuals and/or by grouping known causes of variability and including them in the analysis. Remaining variation can be controlled by randomization.
3. Induced variability resulting from interaction between characteristics of the animal and its environment. Because this source of variability arises from the interaction of two variables, it is by far the most difficult to predict and control and it cannot be measured easily. Example: Two genetically identical male mice, caged together since weaning, will develop a social relationship whereby one predominates over the other. Placed within an enriched environment, these two mice may react completely differently: The dominant animal explores and commandeers “key” microhabitats and accesses resources such as food and water, whereas the subordinate mouse fails to secure a base for itself and apparently spends much of its time withdrawing from contact with the other. There may be no increase in overt aggression, but the difference between the two animals may be much greater than if they were housed together in the same barren environment, which is obviously undesirable.

B. R. Howard, B.V.M.S., M.Sc., M.Ed., Ph.D., M.R.C.V.S., is Director of Animal Welfare and Named Veterinary Surgeon at the Field Laboratories, University of Sheffield, Sheffield, UK.

Variability Introduced by the Experimenter

Variability introduced by the experimenter may arise from two sources: the conduct of procedures involving the animal (e.g., injection, oral dosing, surgical intervention) and lack of precision with measurements. Together these two factors may often be the predominant source of variation in experimental results. Time spent in reflecting and bearing on them and on the training and competence of animal care personnel can have a major impact on the numbers of animals required for a particular study.

Conduct of Procedures

Considerable care should be taken in determining how an experimental procedure is to be conducted to ensure that it is performed in a uniform, precise, and robust manner. For example, when administering substances by injection, a fresh syringe and needle should be used for each animal. Delivery of several administrations by titration from a single filled syringe reduces the precision with which it can be manipulated, and it is more difficult to deliver precise volumes accurately from a wider barrel syringe. Ideally, doses should be determined gravimetrically whereby the weight of the empty syringe is subtracted from its filled weight so that the dose delivered is more accurately known. Good technique is necessary to avoid spillage or inconsistency in the course of administration (e.g., when a needle is dislodged during the course of an intravenous injection). Clean sterile conditions should be maintained to provide consistency of the microbiological environment at the site of injection and to avoid subdermal transport of skin debris, which might interfere with the assimilation of the compound.

Animals should be handled carefully to minimize the stress associated with the procedure, which might otherwise not only influence its metabolism but also affect the rate of uptake due to changes in local blood perfusion rates. Satisfactory performance of even simple procedures should be ensured by an appropriate training program within which competencies are assessed, recorded, and confirmed periodically.

Surgical techniques demand all of the care accorded to human patients, although for the common laboratory species, this need is enhanced by the effects of scale, reflecting the animal's much smaller size. High levels of reproducibility can be extremely difficult to achieve, particularly as natural variations might occur in detailed anatomy. However, procedures should be standardized to the extent possible, and residual variability should be measured so that its influence can be taken into account in subsequent statistical analysis. Such measures might include blood loss (weight of swabs), the weight of tissue removed, and the precise location of intervention in relation to anatomical markers. Even relatively minor variations in technique (e.g., inconsistently tying off a particular blood vessel) can influence

the rate of postsurgical repair and even the eventual pattern of vascular perfusion of a tissue. In the latter case, it is conceivable that such a result could lead to changes in vasomotive control of perfusion under different physiological conditions and could introduce a nonquantifiable variable into the experiment.

Imprecise Measurement

Although many sophisticated instruments are now available and can help with the production of quantitative information, many measurements are inherently imprecise. For example, the size, density, and electrophoretic mobility of spots or bands on a gel can depend on many subtle influences that are beyond the control of the scientist. In addition, scans of optical density can give nonlinear results due to color saturation. Counts of cells of a particular type in a microscope field can be strongly influenced by sampling variation. Colorimetric or fluorimetric measurements can be imprecise, particularly when working near the limits of resolution of the technique. Such variation can seriously reduce the statistical power of an experiment to detect any treatment effects. Apart from ensuring that measurement techniques are appropriate and conducted correctly, the main way of controlling such measurement error is to make multiple observations of the variable of interest.

Control of measurement error is particularly recognized in field observations of natural behavior. In such studies, attempts are often made to achieve consistency by periodically cross-checking the performance of different observers against a consistent test situation. Consistency of interpretation can be challenged by the need to assess the animal's actions subjectively. In such instances, it is important to estimate interobserver variability periodically and to introduce measures to reduce it by training and performance comparisons, thereby maximizing the precision with which the results can be reported. Such activities also provide information that may assist re-evaluation of experimental design and make it easier to compare findings with those of subsequent investigations. Nystrom and colleagues (2001) propose a method by which ethological observations, repeated on a number of successive occasions, can be examined for consistency so as to provide guidance on the optimum number of observational data sessions required for a full study. Alternatively, "variance components" analysis can be used to quantify the within- and between-subject variability. This technique makes it possible to estimate how many repeat observations are needed to attain a given level of precision (Cox 1958).

In many ethological investigations, the key is to determine the most appropriate measures of the behavior of interest. For example, motivational states can be difficult to quantify as they are superimposed on (or rather provide a supporting matrix for) ongoing behavioral patterns. Inferences about motivational states might sometimes be made by free-choice profiling observations in which trained observers conduct initial surveys and generate their own indi-

vidual terminologies to describe the behavior of interest on an analogue scale. This qualitative scale is then used as a semiquantitative measurement to examine behaviors within the test situation of interest. Under rigorous conditions, this approach can yield high inter- and intraobserver agreement (Wemelsfelder et al. 2001).

Despite the attractions of descriptive accounts of behavior as a means of improving our understanding and providing material for hypothesis construction, more traditional approaches sacrifice holistic interpretation in favor of the confidence of being able to apply statistical tests to discreet behavioral components. The more reproducible measurements of behavior can be made, the more precise the possible statistical inferences from them and, it could be argued, the more difficult it may be to interpret these inferences! Robust measures are those most likely to yield similar results when repeated by the same or another observer. They often apply only to small components of the behavior of interest, and although they can be made with great precision, their analytical outcomes rarely describe the behavior of interest, adequately. As in all such cases, a balance is needed between the ease of data handling and its relevance to the hypothesis under examination. If a scientist wishes to explore measurements or assessments of several types of behavior (or any other information such as hematological and clinical biochemical measurements), it is important to realize that the types could be correlated. In such cases, some form of multivariate statistical analysis might be appropriate (Everitt and Dunn 2001).

When behavior patterns are dissected into more quantifiable components, decisions must be made on what these components will be. The occurrence of a particular act may be determined by its frequency, the duration of individual acts of performance or bouts of such acts, and the intensity of performance. All of these factors may vary with time and in relation to each other, and they may also depend on the time and circumstances surrounding a particular trigger event. The collection of such data requires careful observer training and review at intervals to confirm that criteria for scoring have not changed. Despite this, it is often appropriate to regard different observers as independent variables and to include them in any analysis as a check on interobserver consistency.

Even within such controlled conditions, it often remains difficult to establish benchmarks on which all observers can agree. One approach is to dissect the behavior into elements, each of which can be given a binary score (present or absent) over which different observers or the same observer at different times would have a high probability of agreeing. A very simple example of such a scheme is shown in Figure 1. This simple score sheet might be used for preliminary observations on agonistic interactions between two male mice, one of which has been recently added to the cage of the other.

The observer might enter observations at 60-sec intervals onto this chart to score the behavior of the introduced mouse. Observations can be captured by video recording to

Date: Time: Observer: Home male: Test male:

Behavior	Time (min)									
	1	2	3	4	5	6	7	8	9	10
Standing motionless, all feet on floor	✓		✓							
Standing motionless, forefeet raised		✓								
Crouching										
Lying					✓					
Approaching resident				✓						
Retreating from resident										
Movement not directed at resident						✓			✓	
Initiating contact with resident							✓			
Receiving contact from resident								✓		
Mutual contact										
Grooming self										✓
Grooming resident										
Initiate fighting										
Resident initiates fighting										

Figure 1 Sample sheet for scoring agonistic interactions in a simple ethological investigation.

enable the same observer (or different observers) to repeat the analysis to establish the level of inter- and intraobserver concordance. Despite such precautions, score sheets (electronic, manual, or tape based) are often seen as unsatisfactory inasmuch as the subtleties of behavior patterns rarely lend themselves to simple binary assessment. This attempt to generate reproducible data can be further refined by automating data collection using activity meters attached to the animal. This approach is particularly appropriate in free-ranging conditions (Hodgson 1982). Examples of other devices include those based on detection of the animal's weight on different parts of the container floor, interruption of beams of infrared light, disturbance of ultrasound fields, radar, and automated video image processing. Use of such recording devices does not overcome problems with interpretation but instead, enables the scientist to develop highly numerical data that are robust and very suitable for statistical analysis. The use of such devices is particularly favored within regulatory environments, where it is important to ensure the quality of data and where the nature of the hypothesis to be tested can be precisely stated. Despite the advantages of these approaches for improving statistical precision, such data often remain open to different biological interpretation.

An alternative strategy, often adopted by behaviorists as compared with ethologists, is to place the animal in a situation where the behavior of interest can be reliably prompted and the animal's responses robustly recorded. T

and Y mazes, conditioning chambers, and a discriminant learning apparatus are examples of this approach. The animal usually undergoes a period of training after which its performance in the same test is determined under changed biological conditions. The abilities of an animal to discriminate color or sound, or to respond to a range of signals, can be inferred from the results of relatively simple statistical tests. A useful albeit somewhat dated review of approaches to behavioral investigations can be found in the literature (Martin and Bateson 1993).

In whole animal physiological and pharmacological studies, many of the variables inherent in behavioral and ethological observations can be controlled. The strain, sex, age, and weight of the animal are often independent of the study; or the animal can be blocked and included in the statistical analysis or predetermined for these characteristics. Anesthesia isolates an animal from conscious contact with its environment. If necessary, the effects of the anesthetic agent can be taken into account by using a range of different agents and incorporating these into the statistical analysis. The degree of control that can be exerted by the investigator can extend to body temperature, and blood pH/gas composition. The population variance is generally much less than in ethological or behavioral investigations. Even when such studies are carried out on conscious animals after instrumentation (e.g., radiotelemetry for cardiovascular variables in dogs), careful preconditioning of the animals and control of the environment in which measurements are made is important. Under such circumstances, radiotelemetry allows the collection of data from conscious but unrestrained and relatively unstressed animals. In such cases too, data are obtained from an animal over a period of time and their reliability can be estimated. It is particularly important, however, to give prior consideration to the way measurements will be made and analyzed. With care, the use of repeat measurement designs can substantially reduce the numbers of animals required for a particular experimental investigation and often can enhance the relevance of the resulting data with normal physiological conditions.

Experimental designs for traditional pharmacological investigations, particularly in the screening of candidate compounds for pharmacological effects on anesthetized animals, incorporate similar levels of control over biological variability. It is also important to pay attention to the way in which data is collected. There can be no excuse for sloppy practices with regard to the calibration and zeroing of measuring equipment (*before* the experiment commences) or the capture of data. The emphasis here tends to be on the reproducibility of data in the hands of different investigators. As a consequence of the considerable emphasis that regulatory authorities place on confidence in such data, such investigations are often conducted to good laboratory practice (GLP¹) standards. GLP was introduced by the US Food

and Drug Administration in the early 1970s as a means of ensuring the reliability of data submitted for the registration of medicinal compounds in the United States. The concept has since become globally recognized.

The GLP standard of experimentation is predicated on the need for reproducibility of experimental information to provide assurances of the quality of work performed. Each procedure carried out within a GLP investigation must be conducted according to a written protocol described in a "standard operating procedure," which defines precisely the way it is to be carried out. Compliance with this procedure is certified by a system of signatures, and records are periodically audited both by designated individuals at the establishment where the data were generated and by independent quality assurance staff.

One consequence of the introduction of GLP has been a considerable reduction in the variance of measurements taken in scientific investigations, and this reduction reflects greater precision in the way procedures are conducted. Under such conditions, it is more likely that fewer animals are needed to obtain a valid estimate of the likelihood of a compound having a particular pharmacological effect as a result of reducing the amount of "noise" resulting from inconsistency in the conduct of procedures.

When procedures are carried out on isolated cells or organs, the level of control the experimenter can exert over the test conditions is much greater again and more closely approximates that of conventional laboratory chemistry, although biological components of culture media (e.g., bovine serum) may vary between batches and over a period of time. Inherent variability of the cells and tissues, and the way in which they have been prepared may also need to be taken into account; this may include not only breed, strain, and age of the animal from which they were derived, but also its diet, health status, husbandry conditions, and access to environmental contaminants.

Inherent Interanimal Variability

Apart from differences in genetic constitution, many other influences can increase interanimal variability. If these effects can be identified and measured, it may be possible to take account of them using an analysis of covariance.

Source of Animals

Genetic drift ensures that laboratory animals received from different suppliers express differences in their genetic constitution, which may be of minor or major importance. In addition, substantial phenotypic diversity can arise from the different circumstances of the rearing of animals (Crabbe et al. 1999) as well as from genetic differences in founder stock. There is strong evidence for a sensitive postnatal period when young animals develop behavior patterns on which subsequent maintenance behaviors and social inter-

¹Abbreviation used in this article: GLP, good laboratory practice.

actions in adulthood depend (Hol et al. 1999). Even relatively obscure effects (e.g., intrauterine position) may be responsible for environmentally related sex ratio alterations in offspring (Vandenbergh and Huggett 1994).

Gartner (1990) has speculated that additional sources of variability may occur, possibly arising at or before fertilization, but that their nature remains obscure.

No two suppliers use identical care regimes or identical macro- and microenvironments for maintaining animals, and they can be exposed to identical extraneous influences such as photoperiods or microbiological milieu. Although it is regarded as good practice to maintain animals at the establishment where they are to be used for a conditioning period (typically 1 wk) after their arrival, this length of time may not be sufficient for the metabolism or immunology of such animals to reach a steady state. There is therefore a strong reason for factoring the source of animals into the experimental design and for running appropriate controls to check whether adaptation is delayed.

Variability in Animal Care Routines

Relatively subtle variations in routine husbandry practices can have a major influence on the behavior, biochemistry, and physiology of animals. Simply changing to a different animal care technician, while other environmental conditions are held constant, can markedly depress the breeding performance of a colony of mice, perhaps for several weeks. The reason for this depression is rarely clear, and another, although usually much smaller, depression can occur when the original care technician later returns to resume duties in the same room. Presumably these effects on colony performance depend on visual, auditory, or olfactory cues taken from the care provider. It would be prudent to ensure that animals are conditioned to the care provider and to the husbandry conditions in which they will eventually be used for at least 1 wk before an investigation begins (Tuli et al. 1995).

Exposure to Different Environments

Animals are aware of and sensitive to the environment they inhabit. Even subtle changes in their surroundings can have considerable effects on their behavior and physiology. Moreover, the space animals occupy—the microenvironment (the climate within the animal's pen, cage, or even compartment of the cage where nest boxes and other inclusions are provided)—might not correspond to the space in the remainder of the room. For example, the temperature at the air outlet of a typical holding room containing experimental mice can register 3 or 4°C higher than the temperature adjacent to the air inlet, and the relative humidity may be between 5 and 10% higher. There will also be spatial differences in chemical, particulate, and microbiological composition of the air. The position of a rack of cages in

relation to the air inlets and outlets, the vertical position of the cage within the rack, and the presence of any inclusions within the cage (including the degree of filling of the food hopper) all can influence the microenvironment to which the animal is exposed (Clough 1999).

Additionally, mice or rats housed on the top of a multi-tier rack are exposed to higher light intensities than those nearer the bottom. Albino animals are unable to restrict the luminance reaching the retina, which can result in retinal damage, presumably alter pineal function, and may even affect responses to psychoactive drugs (McAllister and Brain 1984). Influences of temperature, relative humidity, and light or sound levels or variation are not always easy to predict, but awareness of them should increase the robustness of experimental design.

The auditory spectrum of rats, mice, and many other species extends for much higher frequencies than in the human. Sounds most likely to influence rodents are those in the low to mid-ultrasonic range (typically 20-40 KHz) (Bjork et al. 2000). It is uncommon for these sounds to be measured within an animal room although they may be generated from a number of sources, including metal sinks, central vacuuming systems, air conditioning plants, and electronic equipment. Because the human ear cannot detect ultrasound, its existence can go unnoticed. Sometimes the first evidence of an acoustic problem can be an increased nervousness of the animals when they are handled, a change in reproductive performance (delayed or suspended conception, smaller or no litters, and/or increased preweaning mortality), or change in growth rate.

The nature of litter material can also affect the metabolic characteristics of animals, including the response to administered compounds. Substrates derived from pine and eucalyptus trees can have greater hepatic enzyme-inducing and cytotoxic effects than materials such as ground corncobs (Potgieter et al. 1995). Socially housed Swiss mice display high levels of immobility in stressful situations; these levels are substantially reduced by the antidepressant drug desipramine, whereas mice housed individually are much less affected (Karolewicz and Paul 2001).

Health Status

Clinical disease leads to the appearance of overt signs of illness. Unless its induction is an integral part of the experimental protocol (e.g., in the development of new therapies), sick animals should not be used for experimental procedures.

Subclinical disease is a different matter because it causes no clinical signs and the experimenter can be unaware of its presence. Illnesses of this type have two major consequences: (1) The disease is likely to affect particular organs, depressing their function and thereby altering the animal's physiological and biochemical state in a way that depends on the nature of the disease process. When a disease alters the function of a particular organ or a particular

biochemical pathway, experimental results are likely to be biased if that function is tested by the procedure. (2) The disease adversely affects an animal's fitness in a broad sense, so that it might become susceptible to influences such as scientific procedures, which otherwise might not have affected it.

A greater problem with infection, whether clinical or subclinical, is that the severity of the illness varies according to the individual animal's fitness and the way the disease spreads through a breeding colony. Within any population, a proportion of individuals can be quite severely affected, some may show no apparent effects, and the great majority experience consequences that are distributed between these two extremes. Variability within the overall population is greatly increased. Most studies on rodents are performed using specific pathogen-free animals, which are free from a range of the more intrusive infective conditions found in the natural rodent population. This practice ensures that this source of variability is minimized. Unfortunately, for many species, specific pathogen-free animals are not available, in which case, particular care must be taken in the selection of animals for use in an experimental procedure. Van Loveren and colleagues (2001) have reviewed the way genetic and environmental factors interact with the immune responsiveness of laboratory animals.

Variability Induced by Interactions Between the Animal and Its Environment

Compounded variability arises from the interplay between inherent differences between animals and variables associated with the experimental protocol or with the way the animals are kept and handled. The complication arises from the tendency of animals to respond physiologically, biochemically, and behaviorally to environmental stress or stressors. The result is that the population variance might be substantially higher than would be predicted if the effects were simply additive. As a consequence, the number of animals required to conduct a study of a given statistical power, and to look for a certain size of treatment effect, can be increased greatly.

One major source of environmental variability arises from initiatives to apply environmental enrichment, either by increasing the complexity of the living space or by group housing, with the consequent development of social hierarchies. Although these initiatives are laudable, it is important to recognize that species (and strains within a species) vary in the way they tolerate group housing. In the wild, hamsters are solitary animals, and they may not adjust to group housing in the laboratory. Similarly, male mice of some strains (e.g., SJL and some sublines of BALB/c and C57BL/6) show high levels of aggression when group housed. However, for the majority of strains of rats and mice, individual housing is in itself a stressful situation, which is only partly ameliorated by the provision of enrichment objects to provide shelter, opportunities for play, and environmental ma-

nipulation. If the experimental design requires that test animals be kept separately, it may sometimes be possible to minimize the consequences of social isolation by providing nonexperimental "companions" of the same species. At the very least, visual and auditory contact should be assured by careful arrangement of cages or pens.

Generally, there is evidence that some strains of mice housed in groups of two, four, or eight per cage are less variable in body weight than those housed singly (Chvedoff et al. 1980). When animals are housed socially, most species establish relatively clear-cut dominance hierarchies: One animal (the alpha) effectively assumes supremacy and exerts a primary call on facilities within the housing perimeter. Other animals adopt subordinate positions relative both to the alpha animal and to others within the group. The result of this hierarchy is that one individual (the omega) represents "the bottom of the heap." This animal may be required to spend a considerable part of its time avoiding conflict with the others.

Haemisch and Gartner (1997) report that in barren cages, groups of three male mice tended to form a stable hierarchy in which two mice appeared equally subservient to the third mouse. However, when environmental enrichment objects were provided, a more traditional linear hierarchy appeared, which was associated with continuing aggression. Aggressive behavior in group-housed male BALB/c mice was found to be manifested least when the animals were housed in small groups of three to five (Van Loo et al. 2001).

During the animals' establishment of a hierarchy, there can be a considerable level of aggression that can cause substantial stress and can markedly influence the behavior, physiology, and biochemistry of the animals involved. After a period of days, hours, or, at most, weeks, when a stable hierarchy is formed, levels of aggression generally decrease markedly and endocrine indicators of stress (such as blood corticosterone concentrations) also decrease. When animals are grouped before the start of a study, it is important to allow the hierarchy to become fully established before the experimental protocol is introduced. Hierarchies within relatively small groups are generally relatively stable, although in larger groups this stability may break down. However, if an animal is removed from the group or becomes unwell, competition can arise and be directed at the creation of a new social structure. Whenever possible, the avoidance of change in the composition of groups is recommended. The performance of procedures on animals within a group can also trigger a period of renewed aggression.

When conflict arises between the provision of husbandry conditions to enhance animal well-being and the consequential increased biological variability that necessitates the use of greater numbers, it is necessary to strive for an ethical balance. Even when groups remain stable, social hierarchy can influence the use made of environmental enrichment strategies, particularly when there is competition for favored areas within the animal enclosure. For example, an alpha rabbit within a group-housed pen often spends

much of its time in a preferred location (e.g., on top of a cage inclusion, when provided). Although it is likely that the Council of Europe will recommend provision of a raised area within cages for rabbits, where social housing is practiced, considerable thought must be given to the types of enrichment inclusions provided to ensure that there is an abundance of “favored” positions. Cage enrichments should be selected for their relative simplicity and uniformity—simple devices to screen off part of an enclosure often prove attractive and effective without increasing competition among cage occupants. It is also important to provide animals with the same enrichment object after a cage has been cleaned as the object that had been present previously. This strategy avoids problems with lingering odors of unfamiliar animals.

Much overt aggression takes place during the hours of darkness, when technical staff are not available to observe it; and its consequences may be apparent the next day in the form of bite wounds and other physical injuries. Even when such injuries do not occur and when behavioral changes are not marked, the consequences of stress on population variability may be considerable. It is not always obvious which animals are experiencing the greater levels of stress—those higher within the dominance hierarchy or those lower. However, it is clear that the stress individuals at different positions within a group hierarchy experience alters their biology in ways that cannot easily be predicted. This stress can vary between different environments and from one experimental situation to another (Mering et al. 2001). In the case of larger animals, it may be relatively simple to determine which individuals are dominant and recessive, and analysis of variance can be used to determine whether the experimental outcomes have been influenced.

Another example of induced variability is the interaction of the animal with experimental routines. As noted above, it is very important to ensure that dosing and sampling are precisely replicated (or at least precisely quantified) so that statistical account can be taken of any differences. Animals respond to different handlers/carers in a way that suggests they are less stressed with some than with others; this condition may be reflected in the apparent level of distress animals experience as a consequence of a procedure. It is also important that if dosing or another procedure is to be performed on a number of consecutive occasions, the same technician should conduct the work, and staff should maintain (i.e., hold constant) the environmental conditions as perceived by the animal, undertaking the procedure at the same time each day to maintain the circadian rhythm.

An additional, easily overlooked factor is the timing of experimental work in relation to that of husbandry procedures. The outcome of a dosing regime that takes place shortly after rodent cages have been cleaned out may not be the same as one in which dosing is conducted shortly before cage cleaning. Moreover, the sequence in which animals are dosed is an important variable that is often overlooked. When animals are removed from their cages, they may vocalize (often within the ultrasonic spectrum and therefore

not audible to humans). Although perhaps not a distress call, this vocalization indicates alarm and is perceived by other animals in the room. If the experimental technique causes a degree of discomfort, the animals on whom the procedure has not yet been conducted become progressively more agitated and may become more difficult to handle. Such animals will have an altered endocrinological status compared with those animals dosed earlier in the sequence.

An effect similar to the one described above is recognized when technicians or scientists are being trained in the proper handling of animals. For example, a group of mice repeatedly handled by relatively inexperienced personnel becomes progressively more difficult to handle as the session proceeds. If an experimental procedure is always carried out in a particular sequence, the experimental results are likely to be biased. Thus, if the control compound is administered first, it can be received by relatively unstressed animals; when the test compound is given subsequently to the other group, their demeanor and physiology have changed and the controls are then not necessarily relevant. If dosing is carried out randomly, then population variability—and consequently the number of animals required to achieve a satisfactory scientific outcome—increases.

If the effects described above are likely to be important, it is possible to alleviate them by using a randomized block design. For example, if four dose levels of a compound are to be given, the first four mice would be randomly assigned to one of the four levels and the injections would be carried out before proceeding to the next four, and so forth.

One of the most obvious ways of dealing with the problem of variability is to ensure appropriate levels of training for those conducting procedures. A successful outcome of training should be judged not only by the competence of the individual conducting the procedure, but also by the responses of the animals while it is carried out. An additional strategy involves acclimatizing animals to reduce levels of alarm they might experience. This technique can be achieved by regular handling and, where appropriate, introducing them to the experimental setting before the study commences. Another technique with the added benefit of assisting animal compliance is to offer a reward immediately after the procedure has been conducted. This technique can involve offering a titbit or comforting the animal to relax it before returning it to its cage.

One very effective way of minimizing the variability that can result from the inexpert conduct of procedures is for institutions to establish central facilities where staff are capable of conducting procedures on behalf of scientists. For example, a typical protocol for mouse immunization schedules involves repeated injections to groups of five to ten BALB/c mice. These animals are inbred, and despite the fact that care is normally taken, it is not at all uncommon for some mice (e.g., 1 or 2 of 10) to develop relatively high polyclonal antibody responses to the antigen and for others (e.g., 3 or 4) to show a very weak response. At the University of Sheffield, we have established a central arrangement for the immunization of mice before the preparation of

monoclonal antibodies by tissue culture techniques. By standardizing the techniques, and particularly by deploying a skilled animal technician who can work competently without stressing the mice, we have never failed in obtaining a sufficiently high immune response in at least one mouse in the smaller group size of four that we have used over a period of 2 yr, sufficient to enable *in vitro* manipulation and the culture to proceed.

References

- Bjork E, Nevalainen T, Hakumaki M, Voipio HM. 2000. R-weighting provides better estimation for rat hearing sensitivity. *Lab Anim* 34:136-144.
- Chvedoff M, Clarke Mr, Faccini JM, Irisari E, Monro AM. 1980. Effects on mice of numbers of animal per cage: An 18-month study (preliminary results). *Arch Toxicol* 4(Suppl):435-438.
- Clough G. 1999. The animal house: Design, equipment and environmental control. In: Poole T, ed. *The UFAW handbook on the care and management of laboratory animals*. 7th ed. Oxford: Blackwell Science. p 97-134.
- Cox DR. 1958. *Planning Experiments*. New York: John Wiley and Sons.
- Crabbe JC, Wahlsten D, Dudek BC. 1999. Genetics of mouse behavior: Interactions with laboratory environment. *Science* 284:1670-1672.
- ECVAM [European Centre for the Validation of Alternative Methods]. 1998. ECVAM Workshop Report 20: Reducing the use of laboratory animals in biomedical research: Problems and possible solutions. Nottingham, England. Reprinted in *ATLA* 26:283-301.
- Everitt BS, Dunn G. 2001. *Applied multivariate data analysis*. 2nd ed. New York: Arnold.
- Gartner K. 1990. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab Anim* 24:71-77.
- Haemisch A, Gartner K. 1997. Effects of cage enrichment on territorial aggression and stress physiology in male laboratory mice. *Acta Physiol Scand* 161(Suppl):73-76.
- Hodgson J. 1982. Ingestive behaviour. In: Leaves JD, ed. *Herbage Intake Handbook*. Maidenhead: British Grassland Society. p 113-139.
- Hol T, Van Den-Berg CL, Van Ree JM, Spruijt BM. 1999. Isolation during the play period in infancy decreases adult social interactions in rats. *Behav Brain Res* 100:91-97.
- Karolewicz B, Paul IA. 2001. Group housing of mice increases immobility and antidepressant sensitivity in the forced swim and tail suspension tests. *Eur J Pharmacol* 415:197-201.
- Martin P, Bateson P. 1993. *Measuring Behaviour*. 2nd ed. Cambridge: Cambridge University Press.
- McAllister KH, Brain PF. 1984. The effects of diazepam upon behaviour in a "standard opponent" test depend upon the lighting conditions employed. *IRCS Med Sci* 12:1113-1114.
- Mering S, Kaliste-Korhonen E, Nevalainen T. 2001. Estimates of appropriate number of rats: Interaction with housing environment. *Lab Anim* 35:80-90.
- Nystrom P, Schapiro S J, Hau J. 2001. Accumulated means analysis: A novel method to determine reliability of behavioral studies using continuous focal sampling. *In Vivo* 15:29-34.
- Potgieter FJ, Torronen R, Wilkes PI. 1995. The *in vitro* enzyme-inducing and cytotoxic properties of South African laboratory animal contact bedding and nesting materials. *Lab Anim* 29:163-171.
- Tuli JS, Smith JA, Morton DB. 1995. Stress measurements in mice after transportation. *Lab Anim* 29:132-138.
- Van Loo PLP, Mol JA, Koolhaas JM, Van Zutphen BFM, Baumans V. 2001. Modulation of aggression in male mice: Influence of group size and cage size. *Physiol Behav* 72:675-683.
- Van Loveren H, Van Amsterdam JGC, Vandebriel RJ, Kimman TG, Rumke HC, Steerenberg PS, Vos JG. 2001. Vaccine-induced antibody responses as parameters of the influence of endogenous and environmental factors. *Environ Health Persp* 109:757-764.
- Vandenbergh JG, Huggett CL. 1994. Mother's prior intrauterine position affects the sex ratio of her offspring in house mice. *Proc Natl Acad Sci USA* 91:11055-11059.
- Wemelsfelder F, Hunter AEA, Mendl MT, Lawrence AB. 2001. Assessing the "whole animal": A free-choice profiling approach. *Anim Behav* 62:209-220.