

# Role of Ancillary Variables in the Design, Analysis, and Interpretation of Animal Experiments

Rose E. Gaines Das

## Abstract

During the course of an experiment using animals, many variables (e.g., age, body weight at several times, food and water consumption, hematology, and clinical biochemistry) and other characteristics are often recorded in addition to the primary response variable(s) specified by the experimenter. These additional variables have an important role in the design and interpretation of the experiment. They may be formally incorporated into the design and/or analysis and thus increase precision and power. However, even if these variables are not incorporated into the primary statistical design or into the formal analysis of the experiment, they may nevertheless be used in an ancillary or exploratory way to provide valuable information about the experiment, as shown by various examples. Used in this way, ancillary variables may improve analysis and interpretation by providing an assessment of the randomization process and an approach to the identification of outliers, lead to the generation of new hypotheses, and increase generality of results or account for differences in results when compared across different experiments. Thus, appropriate use of additional variables may lead to reduction in the number of animals required to achieve the aims of the experiment and may provide additional scientific information as an extra benefit. Unfortunately, this type of information is sometimes effectively discarded because its potential value is not recognized. Guidelines for use of animals include, in addition to the obligation to follow humane procedures, the obligation to use no more animals than necessary. Ethical experimental practice thus requires that all information be properly used and reported.

**Key Words:** blocking; concomitant variable; covariate; exploratory analysis; multivariate analysis; outliers; repeated measures

## Introduction

Good laboratory practice, in some instances reinforced by legal regulations, requires observing and recording a number of items of information about experimental animals in addition to recording the response of primary interest to the experimenter. These observations

are closely linked to the ethics of animal use. The ethical use of animals in science and research is widely discussed, and guidelines for the use of animals are widely available (e.g., Home Office 1995; NRC 1996). Underlying many of these guidelines are the principles of reduction, refinement, and replacement (“3Rs”; Russell and Burch 1959). Properly recorded and used, the required observations of animals should enhance the analysis and interpretation of experiments and hence lead to reduction and possible refinement. Moreover, in the course of obtaining the required observations, marginal additional effort may provide additional information if additional variables are also observed and recorded.

In this article, the term *ancillary variables* is intended to cover the array of variables that may be collected in addition to the defined or primary experimental response variable but that may not be directly related to it. This meaning has similarities to, but should not be confused with, the use of ancillary to refer to a statistic. A statistic is defined as ancillary for the estimation of a parameter of interest if the distribution of the ancillary statistic is not dependent on that parameter (Kalbfleisch 1982). As used here, ancillary variables include variables that are recorded but not used in designing the experiment and are not incorporated into the formal analysis of the primary experimental response variable. The way these variables are used depends on factors such as knowledge of the experimental animals, the intended purpose of the experiment, the design of the experiment, and the range of values of the variables. Some variables may be required as part of the experimental procedure and be intended for use in the statistical analysis of the experiment. Other variables may be collected for use as part of the response (e.g., when the response is specified as the ratio of an organ weight to body weight). A number of variables, if not discarded or ignored, may be useful in various exploratory analyses and thus serve an ancillary role in the analysis and interpretation of the experiment.

An experiment as a whole is characterized by variables that are common to all animals within it (e.g., sex, strain of animal, food supply). Such variables, considered “global,” are relevant primarily for comparisons of different experiments.

Some variables (e.g., size or age) may be incorporated into the experimental design as blocking factors. Other variables, which may be incorporated into the statistical analysis, are referred to as concomitant variables (Fisher 1951), covariates (Sokal and Rohlf 1981), supplementary informa-

Rose E. Gaines Das, Ph.D., is Head of Statistics at the National Institute for Biological Standards and Control, Potters Bar, Herts, UK.

tion (Cochran and Cox 1957), conditional variables or explanatory variables (Elandt-Johnson 1982), and ancillary variables (Zelen 1973). Each of the terms mentioned above has been associated with particular methods of statistical analysis. Of these, the term ancillary variable appears to have been least used with particular statistical methods and hence has been adopted for the more general use defined here.

If variables are not explicitly incorporated into the experimental design or analysis, there is an unfortunate tendency to fail to record and use them fully. Nevertheless, use of these variables may improve analysis and interpretation by providing an assessment of the randomization process and an approach to the identification of outliers, may lead to the generation of new hypotheses, and may increase generality of results or account for differences in results when compared across different experiments. Incomplete records and incomplete use effectively waste this information, which may lead to use of more than the minimum number of animals or to unnecessary repetition of experiments.

It is important to record all variables observed as part of the experimental process and to keep all information about an individual animal together as part of the experimental record. I have seen examples of experimental reports that contain weights for four animals in a cage, together with scores at several times for the same four animals, without any individual identification and with no way of linking weights to individual scores at the different times or of linking scores at the different times to one another. In fact, failure to identify the animal and its associated information severely limits the information available from the experiment. Although results pertaining to the ancillary variables may not contribute directly to interpretation of a particular experiment, they may contribute to explanations of differences between experiments when considered globally. Although the strain of mouse or rat is now usually reported, information about the actual body weights, ages, and housing conditions may be relevant in explaining differences between experiments.

## Ancillary Variables in Experimental Design

Experimental design refers to the way treatments are assigned and applied to the available experimental units. The primary aim of experimental design is to ensure that any differences in the measured response variable have resulted from the applied treatment and not from other, uncontrolled variables. An additional aim is to reduce the variability in the measured response to the extent possible, usually by controlling defined variables, so that the effects of the applied treatments can be assessed more accurately.

The design of experiments in the general context of animal experiments has recently been described by Festing and colleagues (2001b). If the available experimental ani-

mals are all similar to one another, then numbering the animals from one to the maximum available and randomly assigning the treatments may achieve the primary aim. However, when the experimental animals can be clearly distinguished from one another on the basis of known or readily observable characteristics (e.g., age, weight, sex), then failure to incorporate these characteristics into the design may lead to increased variability in the response. One consequence of the increased variability may be that treatments cannot be distinguished.

## Completely Randomized Designs

Completely randomized designs rely on assignment of treatments to experimental animals by a procedure that ensures that each animal is equally likely to receive any treatment. The order in which animals are treated should also be included as part of the randomization process to ensure that all differences among animals and times of treatment and/or measurement of response are spread among all treatment groups with equal probability. Only then is it possible to obtain a reliable estimate of experimental variation and to minimize potential bias (e.g., Fisher, 1951). Paradoxically, the unpredictability of randomization is the best protection against the unpredictability of bias in either direction (Kunz and Oxman 1998). However, if there are large differences among the experimental animals and only a few animals are used in the experiment, then experimental variation may be so large that no differences between treatments can be detected. Moreover, the randomization procedure may itself lead to unacceptable differences between groups if the number of animals is small. For example, with two treatments applied to four mice (two male, two female), a random process might assign one of the two treatments to the two males and the other to the two females.

For completely randomized experiments, the customary formal analysis would not account for any such ancillary variables. However, ancillary variables should not be ignored. In the simple example described above, the report of sex of the animals as randomly assigned would raise questions about whether sex may have influenced the experimental results.

## Randomized Block Designs

Blocking refers to direct manipulations by the experimenter to control one or more independent variables. These independent variables may be characteristics of the experimental animals (e.g., sex or litter), or they may be factors imposed by the nature of the experimental environment (e.g., available space in a cage or on a shelf). The experimenter may divide the available animals into groups that are as closely similar as possible, based on these characteristics. The decision about relevant characteristics for such division de-

depends on the nature of the treatment and the experimenter's knowledge about its likely effects.

For example, animals might be divided into male and female groups, genetically different strains, large and small, or young and old. Even if the experimental animals initially formed a single similar group (e.g., 40 mice of a defined strain, age, and limited range of weight), the housing of the mice may dictate that no more than a specific number (e.g., 10) are kept in a cage. The cage then becomes an important variable inasmuch as mice (or any other animals) kept in the same cage may become more similar to one another in important respects than they are to mice in another cage. More generally, social and environmental effects resulting from group housing configurations can vary depending on animal strain, age, and sex and may affect experimental outcomes.

It may not be possible to administer the treatment and/or measure the response for all experimental animals at about the same time on the same day, in which case the day and time of treatment and/or measurement of response would become important variables likely to affect the result and hence possible blocking factors the experimenter should directly control and account for in analysis of the results. "Blocks" may be "complete" or "incomplete." In the first case, there are at least as many animals within each block as the number of treatments, and each of the treatments is assigned at random to one or more of the animals within each block. Blocks are said to be incomplete when the number of available animals within a block is less than the number of treatments. Descriptions of randomized block designs and methods for their analysis are widely available, and descriptions for incomplete block designs are also described (e.g., Box et al. 1978; Cochran and Cox 1957; Montgomery 2001).

## Other Designs

The Latin square design provides for a double type of blocking restriction (usually characterized as "row" and "column") to be imposed on the experimental animals. For a complex and lengthy experimental procedure, time of day and day of the week might be suitable row and column factors. If an experimenter wishes to control simultaneously the effects of a number of factors (e.g., strain, sex, and age of mouse), factorial experiments might be considered. Numerous designs have also been developed for special situations (e.g., Cochran and Cox 1957; Morris 1999).

## Ancillary Variables Incorporated into Statistical Analysis

To the extent possible, experiments are designed to control the variables likely to affect responses to treatments. Nevertheless, variables frequently exist, which can be measured and which affect the results of the treatment but cannot be

readily controlled. However, if it is possible to "adjust" or "compensate" in some way for these variables, the precision of the response may be increased. Covariate and multivariate analyses provide ways ancillary variables may be formally incorporated into the statistical analysis of experimental data.

## Covariate Analysis

The terms *concomitant* (or *auxiliary*) variables and *covariables* are frequently used to refer to variables that, although measurable, are not readily controllable but may nevertheless affect the response of interest. Fisher (1951) exemplified this with an experiment in which animals at a defined age of treatment were not the same weight. Arbitrary adjustment for weight might be considered (e.g., percentage, rather than absolute, increase in weight). However, it is also possible to use the experimental data to calculate the actual effect of the measured variable—in this case, the covariate "initial weight"—on the measured response. Methods for performing this calculation are described elsewhere (e.g., Sokal and Rohlf 1981). In common with all statistical techniques, adjustments for covariates require assumptions—in this case, about the relations between the covariable and the measured response variable—and may generate questions about interpretation (Cochran 1957; Lord 1960). Such questions are particularly likely if there has not been random assignment of treatments to animals (Miller and Chapman 2001). Thus, careful consideration of the biological situation is required before analysis of covariance is used.

## Repeated Measures

In many experiments, animals are observed and measurements are made daily and, in a number of cases, at more frequent time intervals. Over time, repeated measurements can actually lead to increased precision because measurement errors are "averaged out." Several methods for the analysis of such data are available (Keselman et al. 2001; Senn et al. 2000). In many cases, a simple summary of the measurements may be sufficient (Matthews et al. 1990). For example, gain in weight over the experimental period may sufficiently summarize body weights collected daily. In other cases, the slope of a response line, time to peak response, or area under the curve may be an appropriate summary. However, the complete set of measurements forms part of the experimental results and may provide relevant information when appropriately analyzed.

## Multivariate Analysis

In some animal experiments, many different characters are measured for each animal. For example, 10 or more hematological characters may be measured (Festing et al. 1984).

In experiments involving gene micro arrays, several thousand characters may be measured in each individual. The rapid increase in computing capacity and availability of statistical software over the past 30 yr coupled with the inherently multivariate nature of many phenomena of biological interest has contributed to a substantial increase in the use of multivariate statistical methods. A recent comprehensive overview of multivariate methods with discussion of when and how they should be used is available in the literature (Tinsley and Brown 2000).

Some references to multivariate analysis and/or general linear modeling appear to encourage the simultaneous incorporation of large numbers, or even all, of the variables that have been observed. However, the inclusion of “irrelevant” variables and the exclusion of “relevant” variables may each give rise to problems (Hetherington 2000). Appropriately used, multivariate methods are useful statistical tools (e.g., Festing 1972), but their excessive or inappropriate use can result in misleading conclusions.

## Ancillary Variables in Exploratory Analysis

Several key steps are involved in any experiment. Initially the aims of the experiment must be formulated. This formulation should identify the treatments and experimental units or animals to be used and the response variable(s) to be measured. When the aims have been clearly formulated, the experiment must be designed. At this point a number of variables, both dependent and independent, will be considered and variously taken into account.

For example, a decision may be made to use only male mice. Thus, the experimenter has controlled for the possible ancillary variables species and sex, and these global variables now characterize the experiment as a whole. The age of the mice might be specified within a narrow range, thus directly controlling another possible ancillary variable; or the experimenter may wish to study the effects of the treatment in several different strains of mice, in which case a suitable design might be developed to control this variable. These controlled variables can no longer be used in an exploratory way but have instead been incorporated into the experimental design.

As soon as a design has been developed, an experimental protocol must be prepared. The protocol specifies in detail how the experiment will be carried out. Throughout the development of the design and subsequent protocol preparation, the way in which the statistical analysis will be carried out should also be planned. Among other details, the protocol should clearly specify how all data are to be collected, recorded, and identified. The simplest animal experiments might produce data consisting only of animal identification code, treatment code and the measured response. Even for such limited data, the sequences in which the treatments were applied and in which the responses were measured form an essential part of the experimental record.

For example, if all animals receiving treatment A are

treated first, followed by all animals receiving treatment B, changes in response due to order and/or time of treatment cannot be excluded. Such bias has been shown to occur when animals were paired for similarity and treated in pairs, but one treatment was consistently given before the other (Greenberg 1951).

Unless the experiment is quite small, the animals may be housed in different cages, and the cage identification and location form a further part of the experimental record. However, good laboratory practice, and frequently animal care regulations, virtually always requires the collection of additional ancillary variables (e.g., animal weight). It may be possible to test blood samples required for experimental purposes for additional markers of disease or infection to confirm “pathogen-free” status of animals. The experimental protocol should specify how to record all of this information and should ensure that all data pertaining to each animal are correctly linked and identified.

At the conclusion of the experiment, the records associated with each animal must include the treatment it received and the measured response. In addition, it is important to identify the position in the sequence in which the treatments were applied, the experimenter (if not constant), the cage (if the experiment involved more than one), and any other ancillary information (e.g., weight of animal on arrival in the animal house, weight at commencement of the experiment, weight at conclusion of the experiment, and any other recorded observations about the animal). If any of this information is not available, the experimental protocol may be questioned.

The incorporation of many ancillary variables into the formal analysis of experimental results is neither appropriate nor desirable. Incorporation of irrelevant variables into the analysis or inappropriate incorporation of a possibly relevant variable may lead to confusion. Even when a relation is known to exist between the variables, it may not be relevant in the context of a particular experiment.

For example, it may be known that organ weight is related to body weight in rats. However, if an experiment is carried out using a homogeneous group of male rats initially differing in body weight by  $\leq 10$  g, then calculation of an adjustment for the ancillary variable initial body weight may not be worthwhile because accurate determination of relations between variables requires a wide range of values. For this example, the experimental rats have been selected to have very similar weights, and determination of body weight for a single rat is subject to variation (e.g., depending on whether the rat recently ate or excreted). Thus, any relation of body weight to other variables may not be detectable. Body weight nevertheless remains as an ancillary variable both within the experiment and globally. The need for previous information about the relations between variables if an adjustment is to be attempted has been described (Shirley and Newnham 1984). In the absence of background information, some attempts to adjust may lead to a loss of power in tests about the treatments.

## Use of Ancillary Variables to Detect Outliers

Outliers, or extreme values that have a disproportionate influence on the analysis, may distort the results of any statistical analysis and lead to incorrect conclusions. Both detection and determination of the cause(s) of outliers present statistical problems for which there are no clear-cut solutions (Barnett and Lewis 1994). Some outliers may result from transcription errors (of data recording or entry) and can be corrected. Ancillary variables may provide a means of identifying some outliers with an explanation of their possible cause(s). Initial and final body weights are frequently recorded as part of good laboratory practice. A cause for the outliers (e.g., possible infection) may be indicated if: (1) these variables are appropriately linked in the experimental records, and (2) it can be shown that the one or two apparent outlying responses (e.g., organ weights or biochemical test results) are associated with the one or two animals that also showed weight losses over the course of the experiment—in contrast to all other animals that showed weight gains.

In a recent study (Storring and Gaines Das 2001) in which ovarian weight relative to body weight was measured in rats as response to follicle stimulating hormone, body weights and organ weights were also reported separately for many experiments. A marked outlying response was detected in one experiment in which weights of the 48 female rats were within the range 73 to 90 g with the exception of one, which was reported as 58 g. When the data were analyzed, it could no longer be determined whether this was a recording error; however, the difference was so extreme that it was omitted from analysis.

In Figure 1, data are shown for two hematological variables in mice—red blood cell count (RBC<sup>1</sup>) and hemoglobin content (HGB<sup>1</sup>)—either of which might be considered ancillary to the other, in mice in response to different doses of chloramphenicol succinate (study reported by Festing et al. 2001a; raw data, M. F. Festing, 2001, personal communication). Plots of each variable separately (Figures 1a and b) do not indicate that any one HGB or RBC value differs markedly from other values in the same dose group, although the alert scientist may question the cause of the greater variability of both RBC and HGB values in the group treated with dose 1000 mg/kg and of HGB values in the group treated with dose 2000 mg/kg (Bartlett's test for homogeneity,  $p < 0.01$  for HGB,  $p < 0.1$  for RBC). However, a plot of the two variables in terms of one another (Figure 1c) suggests that one value differs markedly from the common relation between these two variables and, under the assumption of similar relations between these variables in all groups, markedly distorts this relation in the dose group in which it occurs (Figure 1d, correlation coefficient

$r > 0.8$ ,  $p < 0.01$  in all other groups compared with  $r \sim 0.5$ ,  $p > 0.4$  for group with dose 2000).

When “outlying” values such as those described above are detected, they should be considered from both statistical and biological viewpoints. First, laboratory notebooks should be checked to determine whether the value is correct. If there is no evidence of a mistake or transcription error, analysis should be carried out both including and excluding the outlier to learn how it affects the conclusions. If the value is excluded, this exclusion should be explicitly reported and reasons given in the report of the experiment.

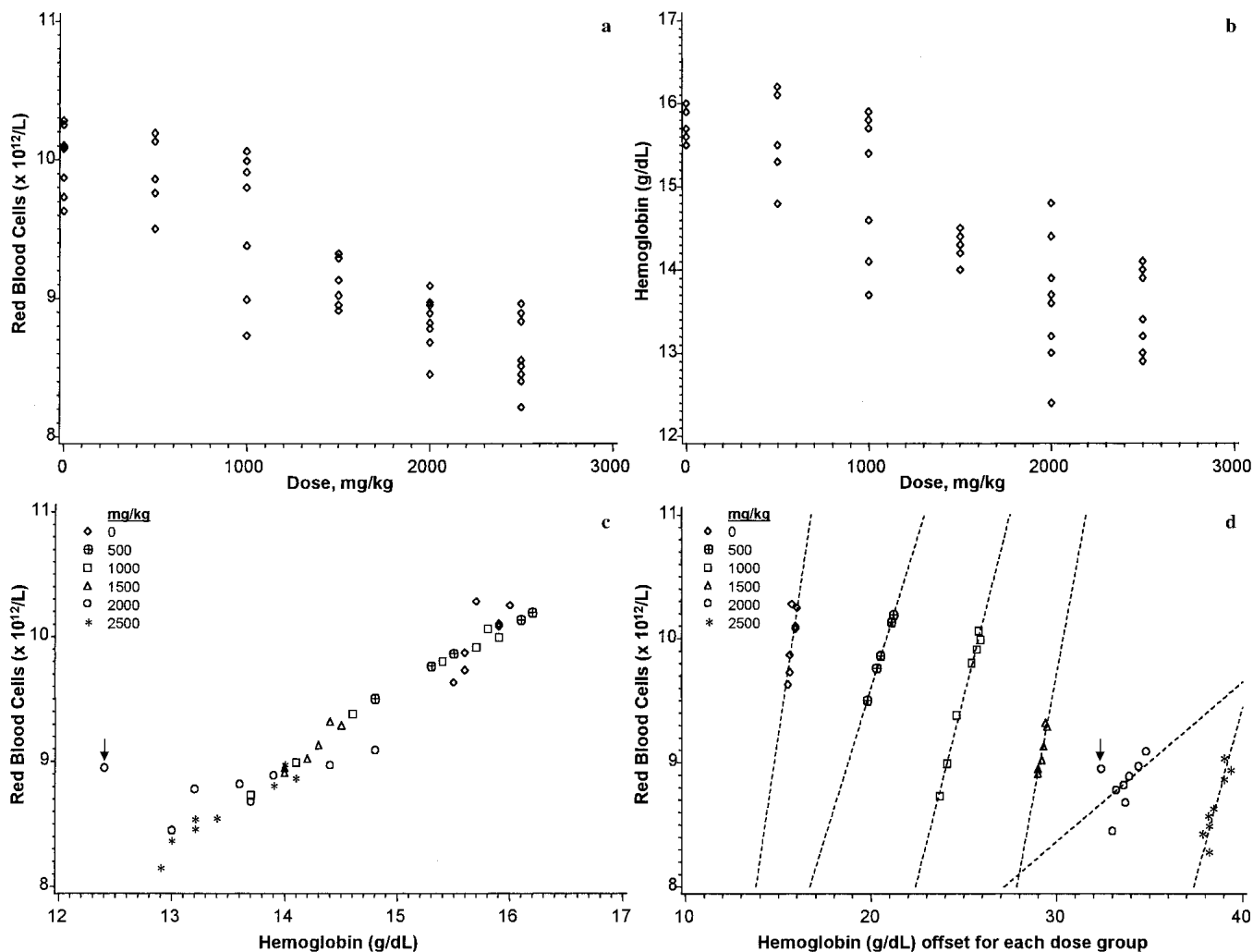
## Use of Ancillary Variables to Assess the Randomization Process

The validity of the interpretation of the results of virtually all statistical techniques is based on the assumption that every observation is statistically independent of every other observation. Small violations of this assumption can seriously affect the validity of any results. The basis for this assumption is frequently the process of random assignment of treatments to animals (e.g., Fisher 1951). Ancillary variables provide one approach to assessing the validity of this assumption. This problem has been considered from a slightly different viewpoint in the clinical literature (e.g., Altman and Dore 1990), where “base-line variables” are used to assess the comparability of groups of randomly assigned patients. The comparability of groups of animals can also be considered. However, it is noted that these tests for comparability also provide an indirect assessment of the “fairness” of the randomization procedure.

One of the likely effects of nonindependence of observations is a reduction in the estimated variance that provides the basis for many statistical tests and, as a consequence of this reduction, greater statistical significance of these tests. The reduction in variability may occur because of the accumulation of nonrandom effects. For example, in many experiments animals are randomly assigned to treatment groups, but the animals that are to receive a specified treatment are then, for convenience, put together in the same cage. The different cages of animals may then be kept in different positions, with more or less light, ventilation, or other environmental factors, and will frequently be treated as a group. This arrangement results in greater similarity for animals within the same cages, and greater differences between cages. Likewise, nonrandom positioning of cages may lead to all cages for one treatment being kept together in sequence followed by all cages for another treatment, with resulting greater similarity among cages with the same treatment. Clearly the variance for more similar animals is smaller than the variance for less similar animals.

If the randomization process has been properly carried out, and no additional nonrandom factors (such as common cages) have been introduced, then the groups of animals should not differ significantly with respect to ancillary variables that are not affected by the treatment. For example, the

<sup>1</sup>Abbreviations used in this article: HGB, hemoglobin content; RBC, red blood cell count.

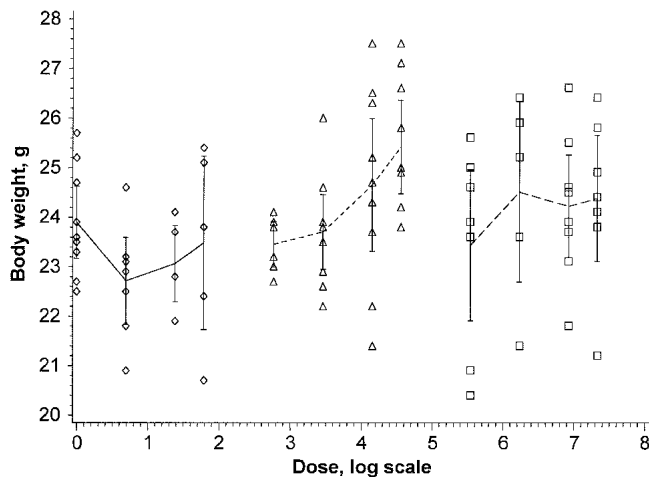


**Figure 1** (a) Red blood cell counts for mice treated with different doses of chloramphenicol succinate. (b) Hemoglobin content of blood for mice treated with different doses of chloramphenicol succinate. (c) Red blood cell counts and hemoglobin content of blood for mice treated with different doses of chloramphenicol succinate (identified by symbols in key). (d) Red blood cell counts and hemoglobin content of blood for mice treated with different doses of chloramphenicol succinate (identified by symbols given in key) with hemoglobin content offset by an additional five for each successive dose. Dashed lines reveal linear regression relation between the two variables within the dose group. Arrow identifies possible outlier in c and d.

initial mean body weights of groups of randomly assigned animals will not be the same, and similarly for other ancillary attributes (e.g., hematocrit). Based on statistical theory, analysis of variance of body weights (or other ancillary attributes) of treatment groups would be expected to show a significant difference with probability  $\leq 0.05$  for 5% of random assignments. Nevertheless, the finding of a significant difference should prompt consideration of the randomization process. If in one experiment individual body weights overlap between groups but mean body weights differ between groups with probability 0.05, this may be a statistical result of the randomization and the process itself may be satisfactory. Simply transferring animals among groups to achieve similar mean weights or “nonsignificant differences” invalidates the randomization process, and such a difference should be noted but is unlikely to require any

further action. In contrast, if such differences are observed in multiple (e.g., 3 or 4) similar experiments, or if the probability is 0.01, the randomization process itself may be unsatisfactory or may have been ignored. Moreover, the factor leading to the differences may also affect the response to the treatments. If body weight (or other ancillary attribute) is not affected by the treatment, then final body weights and changes in weight should also not differ significantly between the groups.

Significant differences in results for ancillary variables have been shown to be associated with significant differences in treatment effects when nonrandom factors are likely (Storring and Gaines Das 1992). The body weights for mice in one assay in that study (Figure 2) differed significantly ( $p < 0.01$ ) between groups treated with two similar preparations, and body weights for mice treated with one



**Figure 2** Body weights for mice treated with different doses of three similar preparations of erythropoietin, each preparation denoted by a different symbol. Doses are shown on a log scale. Error bars reveal  $\pm 1$  standard deviation for each group, centered on the group mean. Approximately equivalent doses were used for each preparation and are offset for clarity.

preparation showed a significant regression ( $p < 0.01$ ) on dose. These differences in body weight are consistent with nonrandom factors. Data may show that cage “scores” are more appropriate than results for the individual animals (Young 1989), or that cage location is a factor that should be included in the analysis (Mantel 1986). Thus information about the ancillary variables may indicate the need to reinterpret significance levels, or to reconsider the definition of the “unit” on which analysis should be based, and to reanalyze the data accordingly.

### Use of Ancillary Variables to Assess Consistency of Observation

Good laboratory practice and animal care regulations require the observation of animals on numerous occasions during the course of an experiment. The results of these observations should be recorded and linked with the experimental records for each animal. The staff who are involved should be identified in each experiment because there may be staff changes from one experiment to another. In some cases, the final observation may be the response measure of interest. However, the previous observations should not be ignored. In the development of an assay for clostridia neurotoxins (Sesardic et al. 1996), the relevant measured response was a cage score for a group of four mice. However, measurements on the individual mice provided useful information about the accuracy and precision of the scoring by individual staff, the consistency of scoring between different staff, and the accumulation of nonrandom effects due to common caging of commonly treated mice. These factors had important implications for the precision and interpreta-

tion of the final response variable. In particular, it could be shown that use of scores for the individual mice would give rise to statistically misleading results. It was also shown that scoring by different members of staff gave consistently different results.

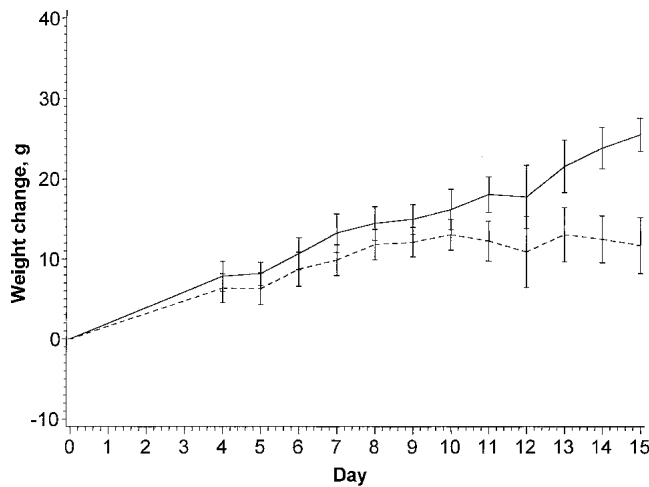
It should be noted that the definition of some of these variables as “ancillary” is dependent on the context and way in which they have been used. Alternate analyses might make different uses of these variables.

### Use of Ancillary Variables to Explain Differences Between Experiments

The conditions of any experiment must be defined and clearly described to permit its independent replication. This description must include details of the animals used and should include at least summaries of all ancillary variables. Full assessment of all factors that may affect biological responses is complex, and failure to provide an adequate description of the factors defining any experiment, to the extent possible, may lead to confusion. Such failure is unethical if it leads to unnecessary repetitions of an experiment. A recent international study (Storring and Gaines Das 2001) revealed the extent to which ancillary variables can assist in accounting for between laboratory differences. This study used an assay method adopted by pharmacopoeias for the control of therapeutic products, and hence a method that might be considered to be well defined. Nevertheless, it was possible to show that differences in weight and age (within the range 19-28 days specified by the European Pharmacopoeia) of animals were associated with significant differences in results from different laboratories. The importance of ancillary information becomes apparent only after such comparisons are carried out.

### Other Uses of Ancillary Variables

In many experiments, the definition of variables as ancillary depends on their use in the context of the experiment. The classical experiment for measuring the effect of growth hormone was based on the increase in body weight of hypophysectomized rats treated for up to 14 days with injections of the hormone. It was not uncommon for rats to be weighed initially and again at the conclusion of the experiment, although it was essential for them to be observed each day. Collection of the weights each day provided additional information about the pattern of growth when growth hormones prepared from bovine or human pituitaries were used (Gaines Das and Rudman 1982). Although the summary measure “total change in weight” was broadly satisfactory for hormones from either source considered separately, the ancillary variables “daily changes in weight” (Figure 3) revealed important differences between hormones from the two different sources in the pattern of weight change. Thus, if a simple summary measure is used for the analysis of



**Figure 3** Mean ( $\pm$  1 standard error) daily changes in weight for two groups, each of 14 hypophysectomized rats, treated with daily injections of nominally equivalent doses of bovine (solid line) or human (dashed line) growth hormone. Weight 0 defined as the mean of weights of the individual rat on the day of first injection and the two preceding days. Day 0 was taken to be the day of first injection. Individual weights were not collected on days 1-3.

“repeated measures” data, then the measures in the complete set may be considered ancillary variables.

Scientists frequently analyze data from toxicological studies based on quantal responses (i.e., proportion of animals showing a defined response at a given time). Records of the time the response occurred and the number of occurrences at a given time provide additional information in the form of time-to-event data. Suitable analysis of these data (e.g., survival analysis and failure-time regression) may provide information about the nature of the biological process. The more efficient use of available information also has the potential for subsequent reduction in animal numbers (Meister 2000). In well-designed and -reported experiments, such data, even if they have not been used in the analysis, will already be available in the form of ancillary variables based on observation of the animals in accord with good experimental practice.

The additional information provided by ancillary variables may suggest a response variable that can be measured earlier or more conveniently than the response selected, and the data may thus be explored for possible “surrogate endpoints” (Buyse and Molenberghs 1998). This process would lead to refinement of the experiment if the animals have a shorter time under the experimental procedures or measurements can be made using less invasive techniques.

## Discussion and Conclusions

To the extent possible, experiments should be designed to control variables that are known or considered likely to influence the response to treatment directly, by blocking or

other suitable design. However, it is not always possible or feasible to control all variables directly. Some of these “uncontrolled” variables may be measured and used in formal multivariate or covariate analysis. Nevertheless, there will remain a large number of variables that are available and frequently recorded as part of good laboratory practice or in conformity with animal care regulations but that are not appropriately included in the formal analysis. Every experimental protocol should ensure that these variables are properly recorded and linked with the animal identification and data records. If this is not done, much useful information may be irretrievably lost, which is both inefficient and unethical.

Complete experimental records should provide clear documentation of the design of the experiment, including the techniques used for randomization. The method of allocation of treatments to the identified animals, linked with all recorded information about the animals, should be described. The selection of the primary variables to be observed and recorded is necessarily guided by the biological theory and aims of the experiment. However, all of the observed and readily available variables should be recorded as part of the experimental record. This documentation requires careful prospective planning and cannot be done retrospectively, however relevant a link may appear to be in retrospect.

Analysis of ancillary variables is unlikely to address the basic scientific theory of a well-designed experiment. Analysis of ancillary variables may nevertheless provide indications of variables that could be relevant in future research. For example, the study by Storing and Gaines Das (2001) suggested that the exact age of the rats used in those experiments may be a more critical variable than had previously been recognized, and that this factor may also be strain dependent to some extent, even though these questions were not directly addressed by that study.

Analysis of ancillary variables has important implications for the interpretation of data because they may contain valuable information about the assumptions that underlie the statistical analysis. Most statistical analysis assumes that the data do not include outliers; and if outliers do occur, then the analysis may be seriously distorted. Unfortunately, indiscriminate omission of data points may also seriously distort any statistical analysis; and omission of any data point, without clear justification, raises questions about the validity of the analysis. Ancillary variables may help to identify and provide possible explanations of outlying data points.

More importantly, however, the value of “statistically significant conclusions” is decreased if these conclusions can plausibly be attributed to irrelevant variables. Thus, if ancillary variables can be shown not to differ significantly between the treatment groups, the plausibility of conclusions may be strengthened. In contrast, ancillary variables that differ significantly between treatment groups require explanation. Decisions in this regard must be informed by an understanding of the biological nature of the complete

experiment. However, these differences may indicate incorrect application of the randomization process or the occurrence of nonrandom factors. Although these differences may not completely invalidate the experiment, they may substantially alter its interpretation.

Valid biological experimentation thus requires the following: correctly recording ancillary variables and linking them with the experimental data; carefully analyzing the ancillary variables; and finally, carefully interpreting the experimental results in the context of the information about the ancillary variables.

## Acknowledgments

The author is grateful to her many colleagues and collaborators at the National Institute for Biological Standards and Control, especially A. Heath, P.L. Storrington, and D. Sesardic, who provided data and helpful discussions; to Issue Editor M.W. Festing, for his thoughtful comments and several apposite references; and to the referees for their constructive comments.

## References

- Altman DG, Dore CJ. 1990. Randomization and base-line comparisons in clinical trials. *Lancet* 335:149-153.
- Barnett V, Lewis T. 1994. *Outliers in Statistical Data*. 3rd ed. New York: John Wiley and Sons.
- Box GEP, Hunter WG, Hunter SJ. 1978. *Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building*. New York: John Wiley and Sons.
- Buyse M, Molenberghs G. 1998. Criteria for the validation of surrogate endpoints in randomised experiments. *Biometrics* 54:1014-1029.
- Cochran WG. 1957. Analysis of covariance: Its nature and uses. *Biometrics* 13:261-281.
- Cochran WG, Cox GM. 1957. *Experimental Designs*. 2nd ed. New York: John Wiley and Sons.
- Elandt-Johnson RC. 1982. Concomitant variables. In: Kotz S, Johnson NL, eds. *Encyclopedia of Statistics*. Vol 2. New York: John Wiley and Sons.
- Festing MF. 1972. Mouse strain identification. *Nature* 238:351-352
- Festing MFW, Diamanti P, Turton JA. 2001a. Strain differences in haematological response to chloramphenicol succinate in mice: Implications for toxicological research. *Food Chem Toxicol* 39:375-383.
- Festing MF, Hawkey CM, Hart MG, Turton JA, Gwynne J, Hicks RM. 1984. Principal components analysis of haematological data from F344 rats with bladder cancer fed N-(ethyl)-all-trans-retinamide. *Food Chem Toxicol* 22:559-572.
- Festing MFW, Overend P, Gaines Das RE, Borja MC, Berdoy M. 2001b. *The Design of Animal Experiments: Reducing the Use of Animals in Research Through Better Experimental Design*. London: Royal Society of Medicine Press Limited.
- Fisher RA. 1951. *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Gaines Das RE, Rudman CG. 1982. Comparison of different measures of growth in the same animals to indicate dissimilarity of growth hormone preparations. In: Gueriguian JL, Bransome ED, Outschoorn AS, eds. *Hormone Drugs*. Rockville: US Pharmacopeial Convention Inc.
- Greenberg BG. 1951. Why randomize? *Biometrics* 7:309-322.
- Hetherington J. 2000. Role of theory and experimental design in multivariate analysis and mathematical modeling. In: Tinsley HEA, Brown SD, eds. *Handbook of Applied Multivariate Statistics and Mathematical Modelling*. London: Academic Press.
- Home Office. 1995. *Code of Practice for the housing and care of animals in designated breeding and supplying establishments*. London: HMSO.
- Kalbfleisch JD. 1982. Ancillary statistics. In: Kotz S, Johnson NL, eds. *Encyclopedia of Statistics*. Vol 1. New York: John Wiley and Sons.
- Keselman HJ, Algina J, Kowalchuk RK. 2001. The analysis of repeated measures designs: A review. *Br J Math Stat Psych* 54:1-20.
- Kunz R, Oxman AD. 1998. The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *Br Med J* 317:1185-1190.
- Lord FM. 1960. Large-sample covariance analysis when the control variable is fallible. *J Am Stat Assoc* 55:307-321.
- Mantel N. 1986. Pancreatic tumors in male rats elicited by corn oil gavage. (Letter). *J Natl Cancer Inst* 77:305-307.
- Matthews JNS, Altman D, Campbell MJ, Royston P. 1990. Analysis of serial measurements in medical research. *Br Med J* 300:230-235.
- Meister R. 2000. Improving data analysis: New aspects in toxicity testing using time-to-event data. In: Balls M, van Zeller AM, Halder ME, eds. *Progress in the Reduction, Refinement and Replacement of Animal Experimentation*. New York: Elsevier Science. p 759-778.
- Miller GA, Chapman JP. 2001. Misunderstanding analysis of covariance. *J Abnorm Psych* 110:40-48.
- Montgomery DC. 2001. *Design and Analysis of Experiments*. 5th ed. New York: John Wiley and Sons Inc.
- Morris TR. 1999. *Experimental Design and Analysis in Animal Sciences*. New York: CABI Publishing.
- NRC [National Research Council]. 1996. *Guide for the Care and Use of Laboratory Animals*. 7th ed. Washington DC: National Academy Press.
- Russell WMS, Burch RL. 1959. *The Principles of Humane Experimental Technique*. Reprint 1992. Wheathampstead: Universities Federation for Animal Welfare.
- Senn S, Stevens L, Chaturvedi N. 2000. Repeated measures in clinical trials: Simple strategies for analysis using summary measures. *Stat Med* 19:861-877.
- Sesardic D, McLellan K, Ekong TAN, Gaines Das RE. 1996. Refinement and validation of an alternative bioassay for potency testing of therapeutic *Botulinum* type A toxin. *Pharmacol Toxicol* 78:283-288.
- Shirley EAC, Newnham P. 1984. The choice between analysis of variance and analysis of covariance with special reference to the analysis of organ weights in toxicological studies. *Stat Med* 3:85-91.
- Sokal RR, Rohlf FJ. 1981. *Biometry, The principles and practice of statistics in biological research*. San Francisco: W.H. Freeman and Company.
- Storrington PL, Gaines Das RE. 1992. The International Standard for Recombinant DNA-derived Erythropoietin: Collaborative study of four recombinant DNA-derived erythropoietins and two highly purified human urinary erythropoietins. *J Endocrinol* 134:459-484.
- Storrington PL, Gaines Das RE. 2001. The Fourth International Standard for Human Urinary FSH and LH: Specificities of LH seminal vesicle weight gain assays in the collaborative study differ between laboratories. *J Endocrinol* 171:119-129.
- Tinsley HEA, Brown SD, eds. 2000. *Handbook of Applied Multivariate Statistics and Mathematical Modelling*. London: Academic Press.
- Young SS. 1989. What is the proper experimental unit for long-term rodent studies? An examination of the NTP benzyl acetate study. *Toxicology* 54:233-239.
- Zelen M. 1973. Keynote address on biostatistics and data retrieval. *Cancer Chemother Rep* 4(Pt 3):31-42.