

Randomization and Analysis of Body Weights in Rodent Toxicity Studies

JEN-FUE MAA

Corning Hazleton Inc., P.O. Box 7545, Madison, Wisconsin 53707

In rodent toxicity studies, one-way analysis of variance (ANOVA) for body weights is commonly applied to a study with or without the assignment to groups of animals stratified by initial body weights. It has been shown that the one-way ANOVA of body weights for a complete randomized design has less efficiency compared with the analysis of two-way ANOVA from a randomized block design or compared with the analysis of covariance from a complete randomized design. The comparison of the efficiency derived from the mean squared errors of a randomized block design with that from a completely randomized design is also discussed. This study includes a discussion of the correlation between growing body weights and initial body weights followed by the statistical analysis of data from different designs.

Key Words: Design, Block, Relative Efficiency.

INTRODUCTION

Body weight stratification for group assignment is frequently applied in a large animal (such as primate or canine) toxicity study because the study design usually has a small number of animals per group and a large body weight variation. Rodent toxicity study designs include randomized block design (RBD) using body weight stratification; completely randomized design (CRD), which assigns animals at random to each group; and a special stratification design. The special stratification design assigns animals to the largest group (i.e., to the group with the largest number of animals) by blocking body weight, then assigning the remaining animals to the second largest group, again by stratification, and repeating the process until all animals are assigned. This design also works if one begins with the assignment to the smallest group and has an algorithm implemented in some medical data base system, but there is no specific statistical model developed for this design. Thus, pharmaceutical companies usually treat the special stratification design as CRD and use one-way ANOVA for the statistical analysis. The baseline body weights (means or standard deviations) from groups from CRD occasionally are not homogeneous. Consequently, analysis of covariance (ANCOVA) is necessary for analyzing the post-treatment body weight data using the baseline data as the covariate. Literature about randomization schemes can be found. For example, CRD with censored criteria at some nominal level, such as $p \leq 0.10$, was cited (Gad and Well, 1994). The complete randomization offers a degree of

simplicity (Arnold, 1991); however, other distribution schemes, including blocking for body weights, littermates, or cages, were considered. Specifications of the National Toxicology Program (NTP, 1992) include examples of body weight stratification using weight classes with all outliers removed. In general, through blocking, the RBD ensures that both means and variations are homogeneous at the beginning of the study. Thus, common sources of variability in an experiment can be systematically controlled (Montgomery, 1984). In this study, we focus on the effect of randomization and the corresponding analysis by comparing the residuals (error) variation when using baseline adjustment or block designs. We discussed the correlation of growing body weights followed by the statistical analysis of data from each design.

METHODS

Animals and Observations. One hundred male and 100 female Crl:CD[®]/BR rats and 100 Crl:CD-1[®](ICR)BR mice were obtained from Charles River Laboratories, Inc. Animals were provided food and water *ad libitum*. Individual body weight and food consumption data were recorded weekly for the first 3 months and monthly thereafter until the animals died or until the 2-year terminal sacrifice. Animals were about 6 weeks old on Day 1 of the study. The correlations between body weights on Day 1 and those at the following intervals were made to illustrate the influence of baseline data on the aged body weights.

Design and Analyses. To show the difference between CRD and RBD, we divided these 100 animals of each sex into four groups to mimic a toxicity study of each design. The mean squared errors were obtained for each design and for each sex by using a one-way ANOVA or a two-way ANOVA separately. Let a and b be the two different designs. The relative efficiency of design a to design b for each analysis was estimated by the following formula (Montgomery, 1984):

$$R = \frac{(df_a + 1)(df_b + 3)}{(df_a + 3)(df_b + 1)} \cdot \frac{\sigma_b^2}{\sigma_a^2} \quad (1)$$

where σ_a^2 and σ_b^2 are the error variances of analyses a and b , and df_a and df_b are the corresponding error degrees of freedom. There are three different analyses to be compared:

- Method A: one-way ANOVA of a complete randomized design
- Method B: one-way ANCOVA of a complete randomized design
- Method C: two-way ANOVA of a randomized block design

Mean squared error of each analysis was used to estimate the error variation. An average of 10 mean squared errors from SAS/STAT® GLM procedures from 10 replicates using freshly generated random

numbers from SAS RANUNI® function and SAS/STAT rank procedure for randomization was reported as the estimated error variation.

RESULTS

For both sexes, Figures 1 and 2 (rats and mice, respectively) show the correlation between aged body weights and Day 1 body weights. The correlation coefficients are high in the first few months with decreasing, yet positive, trend as animals get older. Thus, adjustment for pretreatment baseline data is important when analyzing post-treatment body weights, particularly for short-term toxicity studies. In Figures 3, 4, 5, and 6, the relative efficiency is calculated using (1) where the denominator error variation is estimated from Method A. Numerator error variation is estimated from Methods B or C. Figures 3, 4, 5, and 6 also show that both Methods B and C improved the efficiency of body weight analysis by reducing the error variation. Method B is more efficient than Method C. Because baseline values are used as the covariate in ANCOVA and as the blocking source in RBD, further error variance reduction is limited if both factors are included in the analysis. The component of variation from each model can be demonstrated by the analysis of Day 8 body weights of male rats. Let SST and SS_k be the total sum of squares and the sum of squares from source k , respectively. A single run of the analysis from appropriate SAS GLM procedures of Day 8 body weights for each method produced the results in Table 1.

$$\begin{array}{l} \text{SST} \\ (23008.07, 99) \end{array} = \begin{array}{l} SS_{\text{treat}} \\ (1610.39, 3) \end{array} + \begin{array}{l} SS_{\text{error}} \\ (21397.68, 96) \end{array} \quad (2)$$

$$\begin{array}{l} \text{SST} \\ (23008.07, 99) \end{array} = \begin{array}{l} SS_{\text{treat}} \\ (1610.39, 3) \end{array} + \begin{array}{l} SS_{\text{baseline}} \\ (18921.48, 1) \end{array} + \begin{array}{l} SS_{\text{error}} \\ (2476.20, 95) \end{array} \quad (3)$$

$$\begin{array}{l} \text{SST} \\ (19153.57, 99) \end{array} = \begin{array}{l} SS_{\text{treat}} \\ (374.45, 3) \end{array} + \begin{array}{l} SS_{\text{block}} \\ (5824.41, 24) \end{array} + \begin{array}{l} SS_{\text{error}} \\ (12954.70, 72) \end{array} \quad (4)$$

Table 1 Components of ANOVA models expressed as sum of squares and degrees of freedom

Equations (2), (3), and (4) in Table 1 are derived from Methods A, B, and C, respectively. In this example, the total sum of squares was reduced in RBD compared with that in CRD. In Equation (3), SS_{error} was reduced to 88% of that in Equation (2). This is due to the use of baseline adjustment in the ANCOVA, where SS_{baseline} in Equation (3) is originally part of the SS_{error} in Equation (2) (Cochran and Cox, 1957). In Equation (3), the baseline covariate played an important role in that SS_{baseline} constructed 82% of the total sum of squares. From the relative efficiency formula in

(1), Method B is 8.63 times more efficient than Method A, and Method C is 1.63 times more efficient than Method A when analyzing Day 8 post-treatment body weights. This implies that Method A needs 8.63 times as many animals as Method B or 1.65 times as many animals as Method C to achieve the same degree of efficiency. Figures 3, 4, 5, and 6 demonstrate the improved efficiency of using Methods B and C over using Method A for analyzing body weights measured between Days 8 and 645.

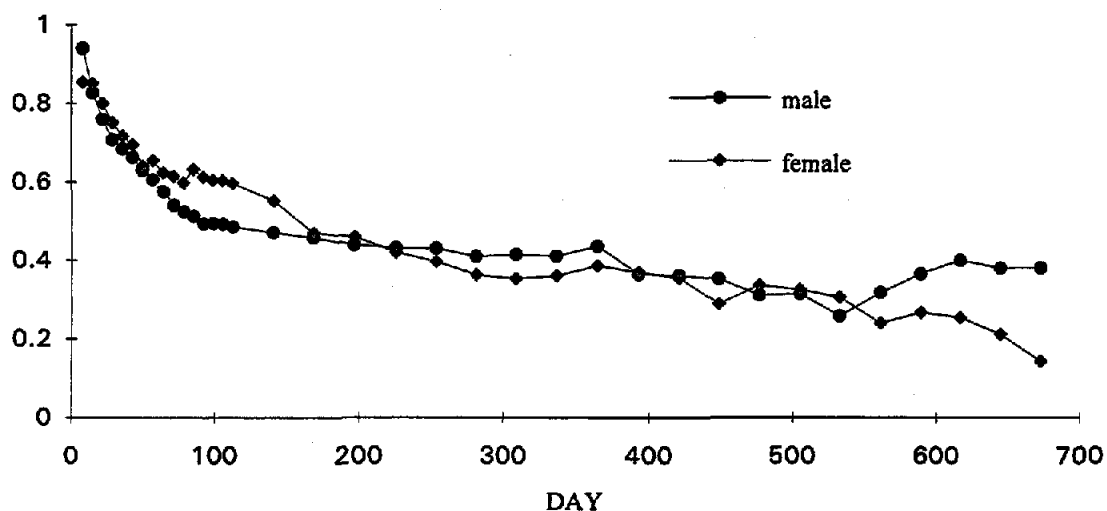


Figure 1 *Correlation coefficient of Day 1 body weights with aged body weights for rats*

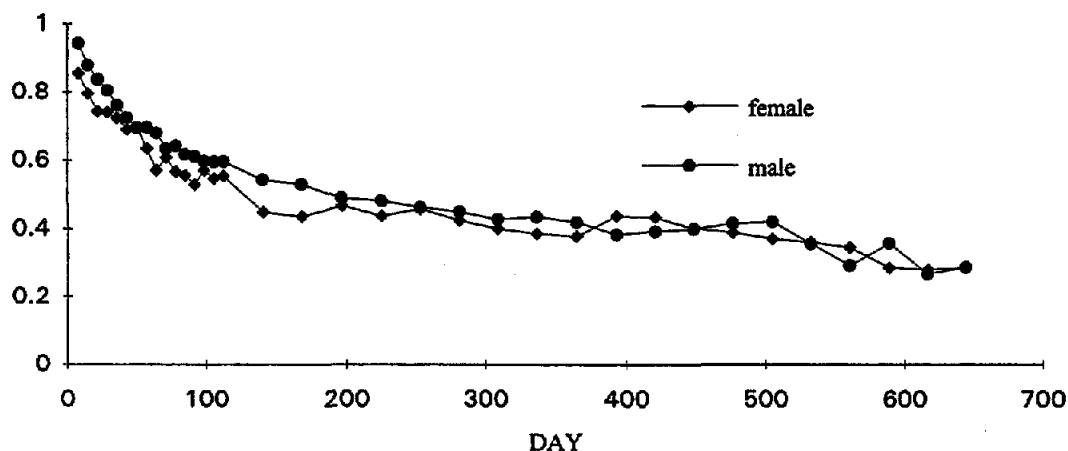


Figure 2 *Correlation coefficient of Day 1 body weights with aged body weights for mice*

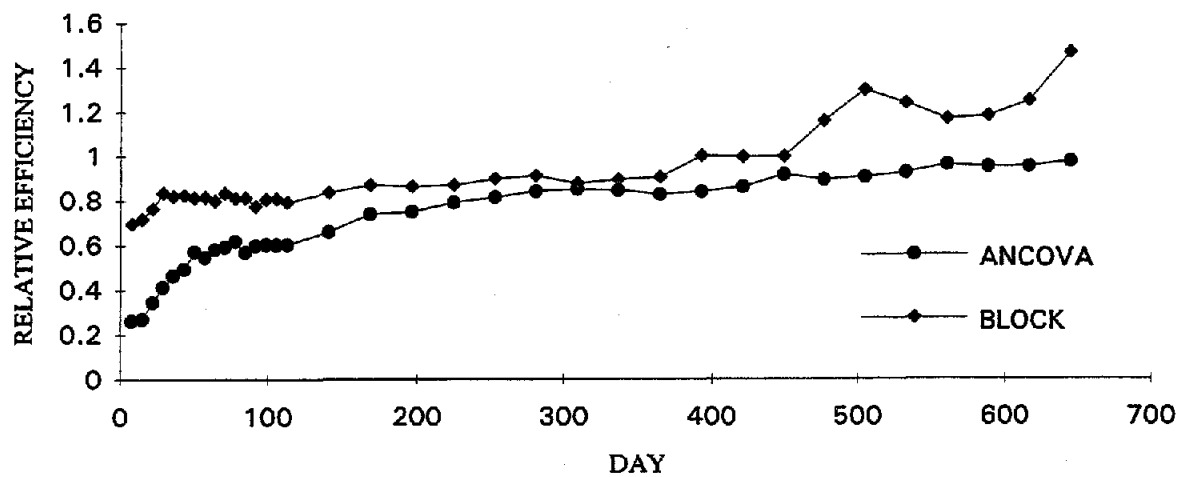


Figure 3 *Relative efficiency of analyses of body weights compared with simple ANOVA for female rats*

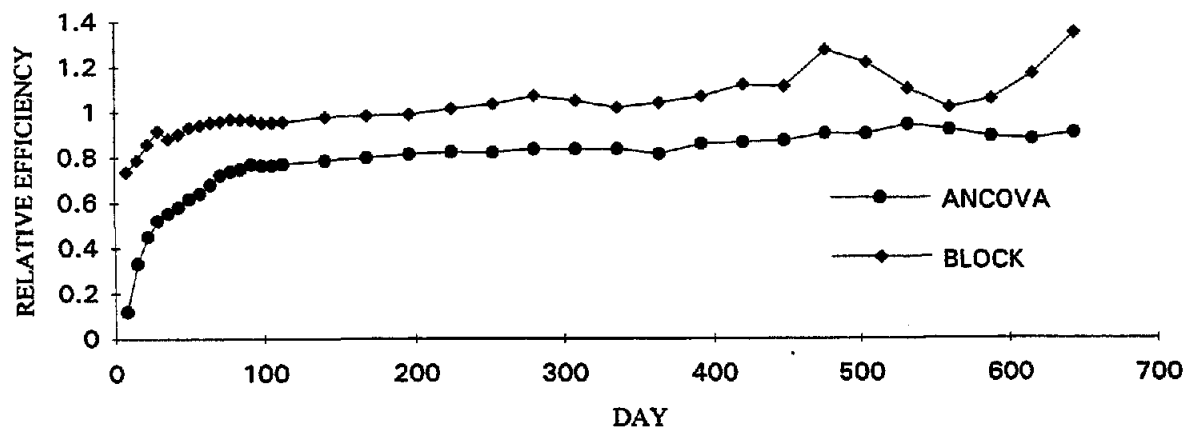


Figure 4 *Relative efficiency of analyses of body weights compared with simple ANOVA for male rats*

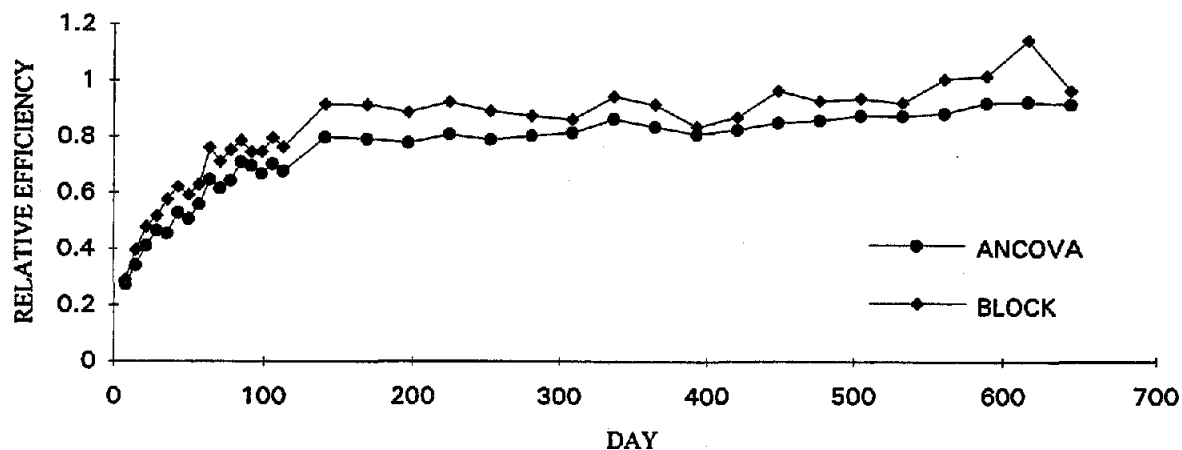


Figure 5 *Relative efficiency of analyses of body weights compared with simple ANOVA for female mice*

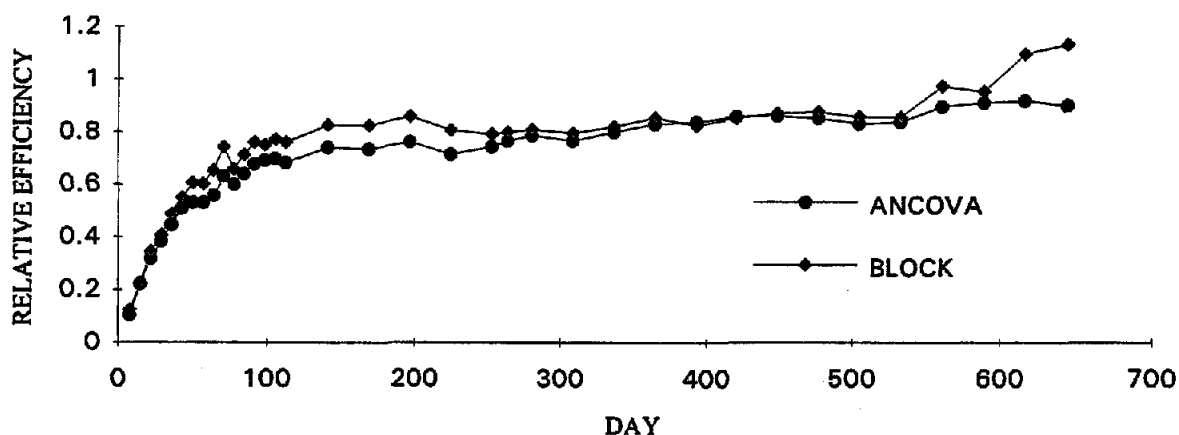


Figure 6 *Relative efficiency of analyses of body weights compared with simple ANOVA for male mice*

DISCUSSION

Body weight is one of the endpoints in toxicity testing. Weil and Gad (1980) used the ratio of mean squared errors and the ratio of F statistics from analyses to compare the efficiency of the covariate when analyzing body, liver, and kidney weights. A 90-day rat study was used to show that a covariate of initial body weights produced a

marked effect in error variation reduction. However, it was relatively ineffective for the analysis of body weight change or for the analysis of the final body weight:initial body weight ratio. Weil and Gad believed that final body weights should be used as the covariate for analyzing liver and kidney weights, but not necessary for the organ-to-body weight percentage data. Other

factors, such as clinical pathology values, may need to be considered. In short, when baseline values are used as the covariate in the analysis, it is more appropriate to report and evaluate the covariate adjusted means rather than the regular means. It has been widespread in preclinical toxicity studies that stratified randomization schemes are used, such as RBD, but a simple one-way ANOVA is applied. This inflates the experimental error, which may result in significant differences in treatment means being undetected. For example, in (4), the sum of squares of error (12954.70), if mistakenly analyzed by a one-way ANOVA, is inflated to 18779.11 ($SS_{\text{block}} + SS_{\text{error}}$). If the animals have been balanced by some other means, the standard procedures will tend to underestimate the normal pattern of variability (Salsburg, 1989). Other analyses, such as applying ANCOVA for CRD only when baseline values are statistically significant, have a drawback in a borderline decision situation of a p-value slightly higher than 0.05, which may seem significant biologically to a toxicologist. Thus, the ANCOVA can be applied without testing the baseline body weights. Comparing Method B with method A, ANCOVA loses one degree of freedom due to the covariate, but greatly improves in efficiency as shown in Figures 3, 4, 5, and 6. Method B is even more efficient than Method C, and both are superior to Method A. There are other appropriate statistical analyses of body weights, such as a repeated measures method, but baseline values should always be included in the analyses. If individual time point analysis is chosen, ANCOVA should be applied when CRD is used. The one-way ANCOVA can also be applied in RBD to achieve better efficiency of the statistical analysis and to ensure that group means and standard deviations are homogeneous at the beginning of a study.

ACKNOWLEDGEMENTS

The author wishes to thank Debra D. Ayres and John Haley at Corning Hazleton Inc. for their assistance in preparing this manuscript.

REFERENCES

Arnold, D. L. (1991), "Subchronic Toxicity Testing," *Statistics in Toxicology*, edited by Krewski, and Franklin, Gordon and Breach Science Publishers, New York, Ch 2.

Cochran, W. G., and Cox, G. M. (1957), *Experimental Designs*, 2nd Ed, John Wiley & Sons, New York, Ch 4.

Gad, S. C., and Weil, C. S. (1994), "Statistics for Toxicologists," *Principle and Methods of Toxicology*, 3rd Ed, Ch 7, edited by Hayes, A. W., Raven Press, New York.

Montgomery, D. C. (1984), *Design & Analysis of Experiments*, John Wiley & Sons, New York.

NTP (1992), Specification for the Conduct of Studies to Evaluate the Toxic and Carcinogenic Potential of Chemical, Biological, and Physical Agents in Laboratory Animals for the National Toxicology Program, unpublished manuscript, pp. 38-40.

Salsburg, D. S. (1989), *Statistics for Toxicologists*, Marcel Dekker, New York, pp. 67-72.

Weil, C. S., and Gad, S. C. (1980), "Application of Methods of Statistical Analysis to Efficient Repeated-Dose Toxicology Tests. 2. Methods for Analysis of Body, Liver, and Kidney Weight Data," *Toxicology and Applied Pharmacology*, 52, 214-226.

SAS and SAS/STAT are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

CrI:CD*BR and CrI:CD-1*(ICR)/BR are registered trademarks of Charles River Laboratories, Inc.

Jen-Fue Maa, PhD
Biostatistician
Corning Hazleton Inc.
P.O. Box 7545, Madison, WI 53707
Phone: (608) 241-7409
Fax: (608) 242-2736
e-mail: 74023.737@compuserve.com