ALTERNATIVE ANALYSES OF CROSSOVER DESIGNS WITH MORE THAN TWO PERIODS

Carla L. Goad

Dallas E. Johnson

Recommended Citation

Alternative Analyses of Crossover Designs with More than Two Periods

Carla L. Goad
Department of Statistics, Oklahoma State University

Dallas E. Johnson
Department of Statistics, Kansas State University

Abstract

A crossover experiment is a special form of a repeated measures experiment. An appropriate analysis of a repeated measures experiment depends on the form of the variance-covariance matrix of the repeated measures. Certain forms of this matrix yield valid analysis of variance F-tests while other forms invalidate these tests. In a crossover experiment where analysis of variance tests are invalid, two alternative tests of a linear contrast of the parameters are proposed. In addition to these approximate t-tests, three alternative methods for testing for equal treatment effects and equal carryover effects are proposed. A simulation study is conducted to evaluate these proposed alternative test procedures for power. Confidence levels and confidence interval lengths are also examined for those procedures from which an estimate of the linear contrast can be made.

Keywords: crossover design, repeated measures

1. Introduction

A crossover design is an experimental design in which each experimental unit (subject) receives a sequence of experimental treatments over time. Sequences usually involve the subject receiving a different treatment during each of the successive experimental periods. However, one or more of the treatments may occur more than once in any sequence. While a completely randomized design could be used to estimate differences between treatments, a crossover design offers some distinct advantages. A crossover design often provides more precision with fewer subjects than a completely randomized design. In crossover designs, each subject serves as its own control. This allows one to estimate the standard errors of estimated differences in treatment effects with a measure of within subject variability which tends to be smaller than between subject variability.

Much has already been written about the advantages, disadvantages, and analyses of two treatment / two period crossover designs. These techniques are nicely summarized by Jones and Kenward (1989) and Ratkowsky, Evans, and Alldredge (1993).

There are two models often used for modeling results obtained from a crossover design. The first is a model that does not include parameters for carryover effects. Such a model is

$$y_{ijk\ell} = \mu + s_i + \delta_{i\ell} + \tau_j + \pi_k + \varepsilon_{ijk\ell} \tag{1.1}$$

for i = 1, 2, ..., s, j = 1, 2, ..., t, k = 1, 2, ..., p, and ℓ = 1, 2, ..., n_i, where $y_{ijk\ell}$ is the response measured on subject ℓ in sequence i receiving treatment j in period k, μ is an overall mean parameter, s_i is the ith sequence effect, $\delta_{i\ell}$ is the experimental error associated with the ℓth subject in the ith sequence, τ_j is the jth treatment effect, π_k is the kth period effect, and $\varepsilon_{ijk\ell}$ is the error associated with the ℓth subject in sequence i that received treatment j in period k. Also it is assumed that $E(\delta_{i\ell}) = E(\varepsilon_{ijk\ell}) = \mathbf{0}$ for all i, j, k, and ℓ, and that all $\delta_{i\ell}$'s and $\varepsilon_{ijk\ell}$'s are independently distributed.

A second model is one that includes carryover parameters. Such a model is

$$y_{ijk\ell} = \mu + s_i + \delta_{i\ell} + \tau_j + \pi_k + \lambda_m + \varepsilon_{ijk\ell} \tag{1.2}$$

where μ, s_i, $\delta_{i\ell}$, τ_j, π_k are defined as in (1.1), and λ_m is the carryover effect of the mth treatment administered in period k-1, where m = 1, 2, ..., t. When k=1, there is no carryover effect. That is, there is no carryover parameter associated with the first period. The model with a parameter for carryover effects presents some other design issues such as confounding of effects and balancing of carryover effects (Cochran and Cox, 1957 and Williams, 1949).

Consider this dairy science example. The effects of three different diet regimens on daily milk production of Holstein dairy cows are to be evaluated. All cows are at a similar point in their lactation cycle when the study begins. Each cow is assigned to a sequence of diet regimens. In each sequence, cows are assigned to the first diet regimen for a two week time period. During the third week of the diet regimen, the average daily milk production (pounds per day) is obtained. The second regimen then follows for a period of two weeks with data being collected during the third week. And the third regimen and measurements follow similarly. A strong period effect is expected since the cows' lactation states change during the study. Although a two week adaptation period to each diet regimen is allotted for each animal, residual effects of diet regimen may still exist. The data for this experiment appear in Table 1 where the three diet regimens are denoted by A, B, and C. Reasonably, the experimenter expects that the production values for each animal to be correlated. Should the usual ANOVA procedures be used in this analysis?

Crossover designs are actually special types of repeated measures designs. Generally, in a repeated measures design each experimental unit is assigned a treatment after which a single response of interest is repeatedly measured over time. A crossover design is more complicated in that each experimental unit experiences a change in treatment from one measurement period to the next.

When analyzing repeated measures designs, correlations between observations taken on a single subject must be considered. Standard analyses of repeated measures designs often use analysis of variance methods appropriate for split plot designs. However, in order for the methods to be appropriate, the variance-covariance structure of the responses taken over time on each subject must have a specific form. Similar requirements are likely to be necessary for crossover experiments that are to be analyzed by analysis of variance methods. When observations are correlated, alternative methods of analysis are needed. Here, alternatives to the usual analysis of variance tests for equal treatment and equal carryover effects are presented and evaluated. Alternative tests and estimates of linear contrasts are also investigated.

Applied Statistics in Agriculture

2. Validity of the Analysis of Variance Tests

As in a repeated measures experiment, a crossover design model can be written in vector form as

$$\mathbf{y}_{i\ell} = \mathbf{X}_i\beta + \varepsilon_{i\ell}, \quad i = 1, 2, \ldots, s, \ell = 1, 2, \ldots, n_i \tag{2.1}$$

where $\mathbf{y}_{i\ell}$ is the px1 vector of repeated measures for subject ℓ in sequence i, $\mathbf{X}_i$ is the appropriate design matrix for sequence i, β is the rx1 vector of parameters where r is the number of parameters specified by (1.1) or (1.2), and $\varepsilon_{i\ell}$ is the px1 vector of random errors associated with subject ℓ in sequence i. Define Σ as the pxp variance-covariance matrix of the error vectors $\varepsilon_{i\ell}$ and let $\mu_i = E(\mathbf{y}_{i\ell})$, i = 1, 2, ..., s. Assume $\varepsilon_{i\ell} \sim$ iid $N(\mathbf{0}, \Sigma)$. Given the assumptions in (1.1) and (1.2), $\Sigma = c\mathbf{I}$ where c is a constant. However, in a repeated measures experiment, correlations among the $\varepsilon_{ijk\ell}$'s are anticipated and accounted for in the analysis.

Under model (2.1) and the assumptions on the error vectors, it can be shown that

$$\hat{\mu}_i = \frac{\sum_{\ell=1}^{n_i} \mathbf{y}_{i\ell}}{n_i}, \quad i = 1, 2, \ldots, s, \quad \text{and} \quad \hat{\Sigma} = \frac{1}{N-s} \sum_{i=1}^{s}\sum_{\ell=1}^{n_i}(\mathbf{y}_{i\ell} - \hat{\mu}_i)(\mathbf{y}_{i\ell} - \hat{\mu}_i)' \tag{2.2}$$

where $N = \sum_{i=1}^{s} n_i$, are sufficient statistics for β and Σ. Furthermore, $\hat{\Sigma}$ is an unbiased estimator of the variance-covariance matrix, Σ, and $\hat{\mu}_i$ is an unbiased estimator of μ_i, i = 1, 2, ..., s. It can be shown that $(N-s)\hat{\Sigma}$ follows a p-variate central Wishart distribution with N-s degrees of freedom and variance-covariance matrix, Σ. That is, $(N-s)\hat{\Sigma} \sim W_p(N-s, \Sigma)$.

The crossover model can be written terms of the sufficient statistics. Define the spx1 vectors $\hat{\mu}$ and μ as

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_s \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_s \end{bmatrix}, \text{ respectively,}$$

and let $\mathbf{H}$ be a spxr matrix such that $\mu = \mathbf{H}\beta$, where r is the number of model parameters in β specified by either model (1.1) or (1.2). A model for the vector of observed sequence means, $\hat{\mu}$, is

$$\hat{\mu} = \mathbf{H}\beta + \varepsilon^*, \quad \text{where } \varepsilon^* \sim N(\mathbf{0}, \Sigma^*) \text{ and} \tag{2.3}$$

$$\Sigma^* = \begin{bmatrix} \frac{1}{n_1}\Sigma & 0 & \cdots & 0 \\ 0 & \frac{1}{n_2}\Sigma & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n_s}\Sigma \end{bmatrix} = \Sigma \otimes \text{Diag}\left(\frac{1}{n_1}, \frac{1}{n_2}, \cdots, \frac{1}{n_s}\right). \tag{2.4}$$

When analysis of variance procedures are used to analyze repeated measures, it is presumed that Σ has Type H structure, which is defined later in this paragraph. In a repeated measures experiment however, Type H structure is not always the case. Consider a crossover experiment where sequences of treatments are defined by mutually orthogonal Latin squares. If there are an equal number of subjects assigned to each sequence, Goad and Johnson (1996) showed that for crossover designs without carryover parameters, the analysis of variance test for equal treatment effects is valid for any non-singular form of Σ. However, when there are an unequal number of subjects per sequence or when carryover parameters are included in the model, the usual analysis of variance tests may not be valid. Goad and Johnson also showed that for the unequal sample size case, valid analysis of variance tests for treatment and carryover exist when Σ has type H structure as defined by Huynh and Feldt (1970). A matrix Σ is said to have type H structure if $\Sigma = \eta\,\mathbf{I_p} + \gamma\,\mathbf{j_p}' + \mathbf{j_p}\gamma'$ for some constant η and some constant vector γ where $\mathbf{j_p}$ is a px1 vector of ones.

Since the analysis of variance tests for equal treatment and/or equal carryover effects are likely to be invalid when Σ does not have type H structure, alternative procedures for a correct analysis must be considered. Some alternative methods for analyzing crossover experiments are proposed and evaluated for power in the next two sections. The alternative methods that are presented depend on the sufficients statistics given in (2.2).

3. Alternative Analyses of Crossover Designs

In this section three alternative approaches to the analysis of crossover designs are suggested for those cases where the variance-covariance matrix does not have type H structure and/or for unequal n_i cases.

3.1 Degrees of Freedom Adjustments

Box (1954a) initially investigated the null distribution of the F-statistic in the analysis of variance under the effect of "group-to-group inequality of variance" in a one-way treatment structure. Box (1954b) continued this investigation by examining the effects of unequal variance in a two-way treatment structure. A randomized block experiment in which the variance changes from one block to the next was one possible scenario Box cited. Box also examined the effects of correlated errors on the null distribution of the analysis of variance F-statistic. Box showed that in the analysis of variance of the two-way treatment structure both the numerator and the denominator degrees of freedom of the F-statistic should be reduced by a multiple of ε where $0 < \varepsilon < 1$, and ε is a measure of how far Σ deviates from compound symmetry. For example, a test for no difference among treatment effects in a crossover design is a within subject test, and

the significance of the test statistic may be affected by the form of Σ. Under H_0: $\tau_1 = \tau_2 = ... = \tau_t$ the usual analysis of variance test statistic would be approximately distributed as $F_{(t-1)\varepsilon,(dfe)\varepsilon}$, where the degrees of freedom for error, dfe, is a function of the total number of subjects, number of treatments, and number of periods. That is, under H_0

$$F = \frac{SSTrt/(t-1)}{SSE/dfe} \; \dot{\sim} \; F_{(t-1)\varepsilon,(dfe)\varepsilon}$$

The reduction factor, ε, developed by Box is given by the following formula

$$\varepsilon = \frac{p^2(\overline{\sigma}_{ii} - \overline{\sigma}_{..})^2}{(p-1)\left(\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}^2 - 2p\sum_{i=1}^{p}\overline{\sigma}_{i.}^2 + p^2\overline{\sigma}_{..}^2\right)} \qquad (3.1)$$

where the σ_{ij}'s are the elements of Σ, $\overline{\sigma}_{ii}$ is the mean of the diagonal terms of Σ, $\overline{\sigma}_{i.}$ is the mean of the terms is the ith row of Σ, and $\overline{\sigma}_{..}$ is the grand mean of the elements of Σ. Note that when Σ possesses compound symmetry, $\varepsilon = 1$. Geisser and Greenhouse (1958) showed that a lower bound for ε is $\varepsilon > \frac{1}{p-1}$, and suggested using this value to yield the most conservative test of H_0 under any existing covariance structure. Greenhouse and Geisser (1959) also used the elements of the sample variance-covariance matrix in (3.1) to compute $\hat{\varepsilon}_1$, an estimate of ε. This correction factor may also be computed by

$$\hat{\varepsilon}_1 = \frac{\left[tr(\mathbf{P}\hat{\Sigma}\mathbf{P}')\right]^2}{(p-1)tr\left[(\mathbf{P}\hat{\Sigma}\mathbf{P}')^2\right]} \qquad \text{or by} \qquad \hat{\varepsilon}_1 = \frac{\left(\sum_{i=1}^{p-1}\lambda_i\right)^2}{(p-1)\sum_{i=1}^{p-1}\lambda_i^2} \qquad (3.2)$$

where $\mathbf{P}$ is any (p-1)xp matrix of orthonormal contrasts and where $\lambda_1, \lambda_2, ..., \lambda_{p-1}$ are the nonzero characteristic roots of $\mathbf{P}\hat{\Sigma}\mathbf{P}'$. Huynh and Feldt (1976) proposed using the reduction factor

$$\hat{\varepsilon}_2 = \frac{N(p-1)\hat{\varepsilon}_1 - 2}{(p-1)(N-s-(p-1)\hat{\varepsilon}_1)} \; . \qquad (3.3)$$

Note that $\hat{\varepsilon}_2$ is not only a function of $\hat{\varepsilon}_1$ but also the total number of subjects, N, and the number of periods, p. Using equations (3.2) and (3.3), it is possible that $\hat{\varepsilon}_1$ or $\hat{\varepsilon}_2$ is greater than 1. In that event, using either method, the estimated correction factor is $\hat{\varepsilon} = \min\{\hat{\varepsilon}_i, 1\}$. Any form of the variance-covariance matrix, Σ, that deviates from type H structure can only decrease the degrees of freedom in determining the significance of F-statistics.

3.2 A Multivariate Approach

To analyze a crossover design using a multivariate approach, consider the definitions based on the sufficient statistics given in (2.3) and (2.4).

Let $\mathbf{a}'\beta$ be an estimable linear function of β. Since $\mu = \mathbf{H}\beta$, $\mathbf{a}'\beta = \mathbf{a}'\mathbf{H}^-\mu = \mathbf{b}'\mu$ where $\mathbf{b} = \mathbf{H}^{-'}\mathbf{a}$ and $\mathbf{H}^-$ is the Moore-Penrose generalized inverse of $\mathbf{H}$. It can be shown that

$\mathbf{b}'\hat{\mu}$ is an unbiased estimator of $\mathbf{a}'\beta$. Under the preceding assumptions, the vector $\mathbf{b}$ can be partitioned as $\mathbf{b} = [\mathbf{b}_1' \ \mathbf{b}_2' \ ... \ \mathbf{b}_s']'$ where each $\mathbf{b}_i$ is a px1 vector. Then

$$\mathbf{b}'\hat{\mu} = \sum_{i=1}^{s} \mathbf{b}_i' \hat{\mu}_i \quad \text{and} \quad var(\mathbf{b}'\hat{\mu}) = \mathbf{b}'\Sigma^*\mathbf{b} = \sum_{i=1}^{s} \frac{1}{n_i} \mathbf{b}_i' \Sigma \mathbf{b}_i \ .$$

Consider the hypothesis H_0: $\mathbf{a}'\beta = a_0$, where a_0 is a specified constant. If Σ is known, a test of H_0 can be based on

$$z = \frac{\mathbf{b}'\hat{\mu} - a_0}{\sqrt{\sum_{i=1}^{s} \frac{1}{n_i} \mathbf{b}_i' \Sigma \mathbf{b}_i}} \tag{3.4}$$

which has a standard normal distribution. When $\hat{\Sigma}$ is substituted for Σ in (3.4), the resulting test statistic does not follow a standard normal distribution. The distributional properties of

$var(\mathbf{b}'\hat{\mu}) = \sum_{i=1}^{s} \frac{1}{n_i} \mathbf{b}_i' \hat{\Sigma} \mathbf{b}_i$ need to be examined before a test statistic can be recommended.

First note that

$$var(\mathbf{b}'\hat{\mu}) = \sum_{i=1}^{s} \frac{1}{n_i} \mathbf{b}_i' \hat{\Sigma} \mathbf{b}_i = tr\left(\sum_{i=1}^{s} \frac{1}{n_i} \mathbf{b}_i' \mathbf{b}_i \hat{\Sigma}\right) = tr(\mathbf{A}\hat{\Sigma}) \tag{3.5}$$

where $\mathbf{A} = \sum_{i=1}^{s} \frac{1}{n_i} \mathbf{b}_i \mathbf{b}_i'$. Also note that $\mathbf{A}$ is a non-negative matrix. Let $a = rank(\mathbf{A}\hat{\Sigma})$. Then

$tr(\mathbf{A}\hat{\Sigma}) = tr(\Gamma\Gamma'\hat{\Sigma}) = tr(\Gamma'\hat{\Sigma}\Gamma)$ where Γ is any pxa matrix of rank a such that $\Gamma\Gamma' = \mathbf{A}$. Since $(N-s)\hat{\Sigma} \sim W_p(N-s, \Sigma)$, it can be shown that $(N-s)(\Gamma'\hat{\Sigma}\Gamma) \sim W_a(N-s, \Gamma'\Sigma\Gamma)$.

Next, define the statistic $U = \dfrac{\upsilon \, tr(\mathbf{A}\hat{\Sigma})}{tr(\mathbf{A}\Sigma)}$ where υ is a constant. Using the results for the trace of a Wishart matrix by Mathai (1980), the first two cumulants of U are $\kappa_1 = \upsilon$ and $\kappa_2 = \dfrac{2\upsilon^2}{N-s}$. Setting these equal to the first two cumulants of the χ^2-distribution with υ degrees of freedom (which correspond to the mean and variance of the χ^2-distribution), an approximation to the value of υ, $\hat{\upsilon}$, can be obtained using Satterthwaite's (1946) approach. Thus the approximate degrees of freedom of U are given by

$$\hat{\upsilon} = \frac{(N-s)[tr(\mathbf{A}\hat{\Sigma})]^2}{tr(\mathbf{A}\hat{\Sigma})^2} \ . \tag{3.6}$$

Thus, the distribution of U can be approximated by a χ^2-distribution, and $U \sim \chi^2(\hat{\upsilon})$. The range of possible values for $\hat{\upsilon}$ is

$$(N-s) \leq \hat{\upsilon} \leq (N-s)a \leq (N-s)p$$

since $\dfrac{[tr(\mathbf{A}\hat{\Sigma})]^2}{tr(\mathbf{A}\hat{\Sigma})^2} \geq 1$ and $[tr(\mathbf{A}\hat{\Sigma})]^2 \leq rank(\mathbf{A}\hat{\Sigma})[tr(\mathbf{A}\hat{\Sigma})^2]$ (Graybill, 1969, p. 303).

With this information, the distribution of the test statistic for H_0: $\mathbf{a'\beta} = a_0$ can be approximated by a t-distribution with $\hat{\upsilon}$ degrees of freedom. That is,

$$\frac{\mathbf{b'}\hat{\mu} - a_0}{\sqrt{\sum_{i=1}^{s} \frac{1}{n_i} \mathbf{b}_i' \hat{\Sigma} \mathbf{b}_i}} = \frac{\mathbf{b'}\hat{\mu} - a_0}{\sqrt{\text{tr}(\mathbf{A}\hat{\Sigma})}} \dot{\sim} t(\hat{v}) \; . \tag{3.7}$$

An approximate (1-α)100% confidence interval for the linear combination $\mathbf{a'\beta} = \mathbf{b'\mu}$, is

$$\mathbf{b'}\hat{\mu} \; \pm \; t_{\alpha/2,\hat{\upsilon}} \; \sqrt{\text{tr}(\mathbf{A}\hat{\Sigma})} \tag{3.8}$$

where $t_{\alpha/2,\hat{\upsilon}}$ is the upper $\alpha/2$ critical point of the t-distribution with $\hat{\upsilon}$ degrees of freedom.

Many data analysts are primarily interested in pairwise comparisons of the treatment effects, $\tau_1 - \tau_2$, $\tau_1 - \tau_3$, ... The preceding results are in no way confined solely to pairwise contrasts. The results are applicable to any within subject contrast including comparisons of treatment effects, period effects, and carryover effects.

Overall tests for equal treatment effects and equal carryover effects are also needed. In a manner similar to that used to develop the approximate t-test given in (3.7) an approximate F-test for treatment effects can be developed. Consider a set of simultaneous contrasts of parameters, $\mathbf{C\beta}$, where $\mathbf{C}$ is a qxr matrix of orthogonal contrasts. Note that rank($\mathbf{C}$) = q and that $\mathbf{C}\hat{\beta}$ is an unbiased estimator of $\mathbf{C\beta}$. Also note that
$$\mathbf{C}\hat{\beta} = \mathbf{CH}^-\hat{\mu} \; \sim \; N(\mathbf{CH}^-\mathbf{H\beta}, \; \mathbf{CH}^-\Sigma * \mathbf{H}^{-\prime}\mathbf{C'}) \; .$$
An approximate F-statistic for the hypothesis H_0: $\mathbf{C\beta} = \mathbf{0}$ can be formed by

$$F = \frac{1}{q} \sum_{j=1}^{q} \frac{(\mathbf{c}_j'\hat{\beta})^2}{\mathbf{b}_j'\hat{\Sigma}^* \mathbf{b}_j} \tag{3.9}$$

where $\mathbf{c}_j'$ is the jth row of $\mathbf{C}$, $\mathbf{b}_j' = \mathbf{c}_j'\mathbf{H}^-$, and $\hat{\Sigma}^*$ estimates Σ^* by

$\hat{\Sigma}^* = \hat{\Sigma} \otimes \text{Diag}\left(\dfrac{1}{n_1}, \dfrac{1}{n_2}, \cdots, \dfrac{1}{n_s}\right)$. Then $F = \dfrac{1}{q} \sum_{j=1}^{q} t_j^2$ where t_j's are the approximate t-statistics developed in (3.7).

The problem here is determining an appropriate number of degrees of freedom to be associated with this F-statistic. Since there are q linearly independent contrasts determining q terms in the summation, the numerator degrees of freedom should be q. However, each term in the summation has a different number of degrees of freedom, $\hat{\upsilon}_j$, associated with it. If each of these approximate t-statistics has a different measure of degrees of freedom associated with it, then what measure of denominator degrees of freedom should be associated with the proposed F-statistic above? One possible choice is to use the value that yields the most conservative test. That is, use $\hat{\upsilon} = \min_j\{\hat{\upsilon}_j\}$. Another choice is to use $\hat{\upsilon} = \max_j\{\hat{\upsilon}_j\}$.

3.3 A Mixed Models Approach

A third approach that can be considered is a mixed models approach or a method of using estimated generalized least squares estimates to construct test statistics. To demonstrate, consider once again the sufficient statistics for a crossover experiment. One has

$$\hat{\mu} = \mathbf{H}\beta + \varepsilon^* , \text{ where } \varepsilon^* \sim N(\mathbf{0}, \Sigma^*).$$

Then a uniformly minimum variance unbiased estimator of an estimable function $\mathbf{a}'\beta$ is

$\mathbf{a}'\hat{\beta}_G$ where $\hat{\beta}_G = (\mathbf{H}'\Sigma^{*-1}\mathbf{H})^-\mathbf{H}'\Sigma^{*-1}\hat{\mu}$. In addition, $\mathbf{a}'\hat{\beta}_G \sim N(\mathbf{a}'\beta, \mathbf{a}'(\mathbf{H}'\Sigma^{*-1}\mathbf{H})^-\mathbf{a})$.

Since Σ^* is not known, $\hat{\beta}_G$ cannot be determined. However, Σ^* can be estimated by

$\hat{\Sigma}^* = \hat{\Sigma} \otimes \mathrm{Diag}\left(\dfrac{1}{n_1}, \dfrac{1}{n_2}, \cdots, \dfrac{1}{n_s}\right)$. Then a mixed model estimator or estimated generalized least

squares estimator of $\mathbf{a}'\beta$ is $\mathbf{a}'\hat{\beta}_{EG}$ where $\hat{\beta}_{EG} = (\mathbf{H}'\hat{\Sigma}^{*-1}\mathbf{H})^-\mathbf{H}'\hat{\Sigma}^{*-1}\hat{\mu}$.

Inferences about $\mathbf{a}'\beta$ can be based on the critical points from the standard normal distribution when the n_i's are sufficiently large. However, when the n_i's are small, the distributional properties of the variance of $\mathbf{a}'\hat{\beta}_{EG}$ cannot be determined making it impossible to construct exact test statistics that may follow a t- or F-distribution. For the simple linear hypothesis, H_0: $\mathbf{a}'\beta = 0$, a reasonable alternative is to consider the test statistic

$$t = \frac{\mathbf{a}'\hat{\beta}_{EG}}{\sqrt{\mathbf{a}'(\mathbf{H}'\hat{\Sigma}^{*-1}\mathbf{H})^-\mathbf{a}}} \tag{3.10}$$

and approximate its distribution by a t-distribution with $\hat{\upsilon}$ degrees of freedom where $\hat{\upsilon}$ is the degrees of freedom estimate given in 3.6. An approximate (1 - α)100% confidence interval estimate of the linear contrast $\mathbf{a}'\beta$ is

$$\mathbf{a}'\hat{\beta}_{EG} \pm t_{\alpha/2,\hat{\upsilon}}\sqrt{\mathbf{a}'(\mathbf{H}'\hat{\Sigma}^{*-1}\mathbf{H})^-\mathbf{a}} . \tag{3.11}$$

An approximate F-test for the hypothesis H_0: $\mathbf{C}\beta = \mathbf{0}$ could also be given by

$$F = \frac{1}{q}\sum_{j=1}^{q}\frac{(\mathbf{c}_j'\hat{\beta}_{EG})^2}{\mathbf{c}_j'(\mathbf{H}'\hat{\Sigma}^{*-1}\mathbf{H})^-\mathbf{c}_j} . \tag{3.12}$$

Then $F = \dfrac{1}{q}\sum_{j=1}^{q}t_j^2$ where t_j's are the approximate t-statistics developed in (3.10). An estimate of the denominator degrees of freedom can be computed in the same manner as in the multivariate approach described in Section 3.2.

The power of each of these three alternative tests for equal treatment effects and equal carryover effects and how they compare to the usual analysis of variance tests are examined in Section 4. For simple linear contrasts, the power of the approximate t-tests, as well as the confidence levels and confidence interval length are also examined.

4. Simulation Results

The proposed alternative methods of analysis and the usual analysis of variance methods were examined in a simulation study. Designs chosen for the simulation study were Williams' designs since the treatment and carryover effects are not confounded in these designs. Two crossover designs were examined: p=t=3 Williams design (MOLS, Table 2) and p=t=4 Williams design (Table 3). In order to examine the reliability of these alternative methods, several

covariance structures for Σ were examined. For the p=t=3 designs, models with and without carryover parameters were investigated. In this simulation study, tests for equal treatment effects and equal carryover effects were investigated.

4.1 Methods Used in Simulation

For the vector form of the model (2.1), the sufficient statistics, $\hat{\mu}_i$ and $\hat{\Sigma}$, were simulated for each crossover design. The proposed alternative methods of analysis for simultaneous and single degree of freedom contrasts were compared to the usual analysis of variance F-tests and t-tests for a linear contrast. Tests of treatment effects and of the carryover effects were analyzed.

Variance-covariance matrices having compound symmetry ($\varepsilon = 1.00$) and matrices that deviate from compound symmetry were used in the simulation study. Box's ε was used to characterize a matrix since ε is not a function of the number of subjects per sequence. In this particular study, the variances, σ_k^2, of the observations during each period were chosen such that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2 = \sigma^2$. Without loss of generality, $\sigma^2 = 1$ was chosen.

Let the elements of Σ be given by

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & & \rho_{1t} \\ \rho_{12} & 1 & \rho_{23} & \cdots & & \rho_{2t} \\ \rho_{13} & \rho_{23} & 1 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \rho_{(t-1)t} \\ \rho_{1t} & \rho_{2t} & \cdots & & \rho_{(t-1)t} & 1 \end{bmatrix}.$$

The values of ρ_{ij} were chosen so that a range of values for ε resulted. For a repeated measures experiment, reasonable values of ρ_{ij} are $1 \geq \rho_{ij} \geq \rho_{i(j+1)} \geq 0$ where $1 \leq i < j \leq t$. That is, observations from adjacent periods should be more strongly, positively correlated than those from non-adjacent periods. For this simulation study, the values of ρ_{ij} used are given in Table 4 and Table 5.

One of the hypothesis tests to be evaluated was the test for equal treatment effects, $H_0: \tau_A = \tau_B = \ldots = \tau_t$. When $\Sigma = \sigma^2 I$, the test statistic for testing this hypothesis, $SSTrt/\sigma^2$, has a non-central χ^2–distribution with t-1 degrees of freedom and non-centrality parameter $\lambda = \frac{1}{2\sigma^2}(C\hat{\beta})'[C(X'X)^- C']^{-1}(C\hat{\beta})$ where C is an appropriate contrast matrix for testing H_0 and

X is the design matrix. The power of the test for equal treatment effects depends only on the τ_j's through λ. Now for t treatments, s sequences and n subjects per sequence

$$\lambda = \frac{sn}{2\sigma_\varepsilon^2} \sum_{j=1}^{t} (\tau_j - \bar{\tau}.)^2 \quad \text{where} \quad \bar{\tau}. = \frac{1}{t} \sum_{j=1}^{t} \tau_j .$$

Without loss of generality the values of λ were changed by letting $\tau_A = 0, 0.1, 0.2, \ldots, 1.2$ in model (1.1) while keeping the remaining τ_j's = 0. When $\tau_A = 0$, H_0 is true, and an estimate of test size is obtained. When $\tau_A = 1.2$, the power of the tests neared 1, and larger values for τ_A were not necessary.

In the case of the linear contrasts of the parameters, denoted by $\mathbf{a'\beta}$, the powers of the standard normal test and the t-test are functions of the actual value of $\mathbf{a'\beta}$, a scalar value. To approximate the powers of the test procedures for H_0: $\mathbf{a'\beta} = \mathbf{0}$, $\mathbf{a'\beta}$ needed to assume a value of 0 (to estimate the sizes of the tests) and values of increasing distance from zero to estimate the power of the tests. Since the standard normal and t-distributions are symmetric about zero, only increasing positive values of $\mathbf{a'\beta}$ were necessary. The values of $\mathbf{a'\beta}$ in the simulations were determined by the values of the treatment parameters. In this simulation, the linear contrast hypothesis was H_0: $\tau_A - \tau_B = 0$. As in the case of the analysis of variance F-test for equal treatment effects, the actual values of the τ_j's do not determine the powers of the tests. A range of values for $\mathbf{a'\beta}$ was obtained by using the same method of defining the τ_j's as before. That is, for all simulations τ_A assumed the values 0, 0.1, 0.2, ..., 1.2 while $\tau_B = 0$ always. In the discussion that follows, the value of τ_A is referred to as the "treatment difference."

Tests for equal carryover effects and linear contrasts of the carryover parameters were also simulated. The values of the treatment and carryover differences varied from 0 to 1 by 0.2 in model (1.2). The value of τ_A is again the "treatment difference," and the value of λ_A is referred to as the "carryover difference," where τ_A and λ_A are parameters in model (1.2).

For (1.1) samples of equal size were simulated and for (1.2) both equal and unequal sample sizes were investigated. Tables 6 and 7 provide the conditions of the simulations. The six and four digit numbers in the tables refer to the sample sizes that were used for each sequence. For example, in the p=t=3 Williams' design 233233 means that sequences 1 and 4 (defined in Table 2) had samples of size 2 while all other sequences had samples of size 3. There were 1000 simulations of the sufficient statistics for each τ_A, sample size configuration, and Σ combination for the model with no carryover parameters (1.1) For the model with carryover parameters (1.2) there were 1000 simulations of the sufficient statistics for each treatment difference, carryover difference, sample size configuration and Σ combination.

First, the p=t=3 Williams' design was investigated under the no carryover parameter model (1.1) with varying values of ε. For this design the usual ANOVA F-test for equal treatment effects is valid for all variance-covariance structures when the sample sizes are equal. Alternative test methods were judged to be acceptable if their size and power were similar to the ANOVA. Five different variance-covariance matrices for each equal sample size selection were used in the simulations. The p=t=4 Williams' design was investigated using model (1.1) with varying Σ since the ANOVA F-test for equal treatment effects is not valid for this design unless Σ has type H structure.

The model containing carryover parameters (1.2) was investigated using both equal and unequal sample sizes at two values of ε for the p=t=3 Williams' design. When $\varepsilon = 1.00$, the ANOVA F-test is valid for both the equal and unequal sample size cases. When ε is small the ANOVA F-tests for equal treatment effects and equal carryover effects are invalid. Only the extreme values of $\varepsilon = 1.00$ and $\varepsilon = 0.5586$ are used in the simulations for the model containing carryover parameters.

Since a Williams' design does not confound treatment and carryover effects, the power associated with each of the tests for equal treatment effects was pooled across the values of carryover difference (0, 0.2, 0.4, 0.6, 0.8, 1.0). Therefore, at each level of treatment difference the

power of the test is based on 6,000 simulations. Likewise, the tests for equal carryover effects pooled information from all levels of treatment difference (0, 0.2, 0.4, 0.6, 0.8, 1.0).

4.2. Approximate F-tests

Several methods for testing a hypothesis of equal treatment effects were evaluated for each set of simulated data. The first four were based on the usual analysis of variance F-statistic for testing equal treatment effects. This F-statistic was compared to critical points from the F-distribution with four different choices for degrees of freedom. One choice used degrees of freedom equal to those that would be appropriate when Σ has type H structure. The three others used degrees of freedom adjustments: the lower bound for Box's ε, Greenhouse and Geisser's $\hat{\varepsilon}_1$, and Huynh and Feldt's $\hat{\varepsilon}_2$.

Tests based on the multivariate approach described in Section 3.2 were also evaluated. In this case t-1 test statistics given by (3.7), and an approximate F-statistic given by (3.9) are computed. The approximate degrees of freedom for each of the approximate t-statistics were computed by (3.6). Since the t-1 test statistics do not yield the same value for $\hat{v}_i$, the approximate degrees of freedom for the denominator, alternative degrees of freedom choices were examined. The first two were: $\min_i\{\hat{v}_i\}$ and $\max_i\{\hat{v}_i\}$. A third alternative was a lower bound, N - s, for all $\hat{v}_i$. The p=t=3 Williams' design has $\hat{v}_i$ values that are bounded above by 2(N - 6) = [Rank($A\Sigma$)](N - 6). The p=t=4 Williams' design has $\hat{v}_i$ that are bounded above by 2(N - 4) = [Rank($A\Sigma$)](N - 4) and 3(N - 4) = [Rank($A\Sigma$)](N - 4) depending on the orthogonal linear contrast. The fourth and fifth alternatives were 2(N - s) and 3(N - s). The sixth alternative was the degrees of freedom due to error that is computed in the usual analysis of variance procedure.

For the mixed models approach the approximate F-statistic given by (3.12) is evaluated. The same six denominator degrees of freedom estimates as described for the multivariate approach were used in the mixed models approach. For each of the multivariate and mixed models approaches six observed significance levels were determined according to the six F(t-1,df) distributions where df was determined by each of the above six methods discussed. The power of each of these procedures was obtained.

4.3 Approximate F-test Results

A guide to the notation used in simulation results for the approximate F-tests is summarized in Table 8. All tests were conducted at the nominal 0.05 level.

In the three treatment / three period design, $\hat{v}$ is invariant with respect to the orthogonal contrasts, and rank($A\Sigma$) = 2 (Goad 1994). Hence, F1 and F7 use the unique value of $\hat{v}$.

Among each of the alternative procedures for testing for equal treatment effects, a preferred test is recommended. Criteria for selecting the preferred test are:

1. the test exhibits size and power similar to the ANOVA F when ε = 1.00,
2. the size of the test is conservative across the values of ε,
3. the test demonstrates strong power.

After determining the preferred test from each of the alternative procedures for each of the models, the usual ANOVA test for equal treatment effects is compared to each of the preferred tests from the alternative procedures.

Among the degrees of freedom adjustment techniques, HF is the preferred test across all the values for n and ε when there is no carryover. The HF test was conservative (size < 0.05) in this crossover setting, but it also exhibited a greater power than the other two methods, BOX and GG. The power of the HF test did coincide with the usual analysis of variance test, F, when ε = 1.00, as expected. These statements are true for both of the p=t=3 and p=t=4 cases.

When there is carryover, GG and HF tests for equal treatment effects tended to yield sizes that were larger than 0.05 in both the equal and unequal sample size cases making these tests unacceptable. For testing equal treatment effects, BOX exhibited a test size slightly smaller than 0.05 for all sample size cases when ε was low and was the preferred degrees of freedom adjustment test for equal treatment effects for this reason. For the tests of equal carryover effects, the degrees of freedom adjustment techniques when ε < 1 exhibited test sizes considerably smaller than 0.05. All of the degrees of freedom adjustment techniques were quite conservative in the tests for equal carryover effects. Among the degrees of freedom adjustment methods, HF was the preferred method for testing for equal carryover effects.

For the multivariate methods, first consider designs without carryover. All but F3 performed similarly to F when ε = 1.00. However, F5 and F6 had inflated test sizes for smaller values of ε making them unacceptable. For larger values of ε, each of the other five methods tended to have similar powers. F4 exhibited better power for low values of ε and is the preferred test among the multivariate alternatives. In the model with carryover parameters (1.2), F1 and F4 demonstrate size and power most like F when ε = 1.00 when testing for equal treatment effects. For a low value of ε, F1 is the preferred test for both equal and unequal sample size cases. When testing for equal carryover effects, F1 is again the preferred test of the multivariate approaches.

In the mixed models approach for designs without carryover, none of the tests for equal treatment effects can be recommended due to the fact that the sizes of these tests were at least twice the nominal value of 0.05. F9 was the best of those considered. For designs with carryover, F9 was again the best choice although the size of the test for equal treatment effects was still above 0.05. In the test for equal carryover effects, F9 is the best option when testing for equal carryover effects for all of the sample size cases considered. For both models, all of the mixed models approaches exhibited a rapid increase in power in addition to the large size. When ε < 1, the difference between the power curves associated with F9 and the power curves associated with F, any of the degrees of freedom adjustment, or multivariate approaches is greater than the differences between the sizes of the tests. When sample sizes were small (2 or 3), the mixed models approaches performed poorly with respect to test size.

Figures 1 and 2 give the power curves for the preferred tests of equal treatment effects (F, HF, F4, F9) from each alternative procedure for two values of ε in the four treatment / four period design when the model with no carryover parameters (1.1) is used. These curves are for tests based on three subjects per sequence. For larger values of n and ε < 1, the power curves tended to be similar to those in Figure 2. (Since ε = 1.00 in Figure 2, the power curves for the usual ANOVA and Huynh-Feldt methods coincide.)

Figure 3 demonstrates the power curves for the preferred tests of equal treatment effects (F, HF, F1, F9) from each alternative approach when a model with a carryover parameters is used. Only F1 meets the requirement of test size smaller than 0.05 consistently for all of the sample size choices.

Figure 4 contains the power curves for the preferred tests (F, HF, F1, F9) of equal carryover effects. Of those that satisfy the 0.05 test size requirement, the usual ANOVA F-test has marginally better power than the others. The mixed models approach does not meet the size requirement.

4.4 Tests of a Linear Contrast

For each of the crossover design models (1.1) and (1.2), ten different methods for testing a linear contrast (pairwise difference) of the treatment effects were evaluated for each set of simulated data. For the model containing carryover parameters (1.2), tests of a linear contrast of the carryover parameters were also evaluated. Again, all of these tests were conducted at the nominal 0.05 level. Since a known variance-covariance matrix, Σ, was used in all simulations, the standard normal test statistic given in (3.4) was computed and used as an optimal test statistic.

The usual t-test,

$$t = \frac{\mathbf{b}'\hat{\mu}}{\sqrt{\mathrm{MSE}\sum_{i=1}^{s}\frac{\mathbf{b}_i'\mathbf{b}_i}{n_i}}} \text{ , where } \mathbf{b}'_i \text{ are the } 1\mathrm{xp} \text{ elements of } \mathbf{b}' = [\ \mathbf{b}'_1\ \mathbf{b}'_2 \ldots \mathbf{b}'_s]\ ,$$

was also computed and compared to critical values from the t-distribution with degrees of freedom equal to that of the mean square error (MSE) from the usual analysis of variance.

Four methods were based on a multivariate approach. The test statistic was computed according to (3.7). This test statistic was compared to critical values from the t-distribution using four different estimates of its degrees of freedom. The four degrees of freedom estimates used in the simulation were $\hat{v}$ as computed by (3.6), N - s which is the lower bound for $\hat{v}$, 2(N - s) which is the upper bound for $\hat{v}$ (as determined by the linear contrast τ_A - τ_B), and the error degrees of freedom from the analysis of variance.

4.5 Approximate t-test Results

Table 9 provides a guide to the notation used in the linear contrast power analysis.

In the power analysis the power curves of TU and Z converge to one another as the n_i's become large. As in Section 4.3 a preferred test from each of the alternative approaches is obtained. Then each of these preferred tests is compared to TU and Z with respect to power, confidence level, and confidence interval length.

Among the multivariate alternatives T3 and T4 tend to converge as n increases for both crossover models. When ε is small, T3 is a conservative test. Nevertheless, it has better power than the other methods. Among the mixed models alternatives, T6 is most similar to TU at ε = 1.00. However, all of the mixed models alternatives have inflated sizes. This situation improves as n increases or as ε nears 1. For the model with no carryover parameters (1.1), T3 and T6 are the preferred tests. Figure 5 contains power curves comparing best methods from each of the

four approaches (Z, TU, T3, T6). The multivariate approach (T3) compares most favorably to Z and meets the 0.05 size requirement.

For the model with carryover parameters (1.2), consider first the linear contrast of the treatment effects. T2 tended to control for size better than the other multivariate approach methods, especially when there are large differences among sample sizes. T6 performed better than each of the other three mixed model approaches. For tests of a linear contrast of the carryover parameters, the conservative tests T2 and T6 were again the best. Inflated test sizes occur for the usual t-test for treatment effects, but for the tests of a linear contrast of the carryover parameters the usual t-test does control test size. In both cases the mixed models approach exhibits a test size that is greater than 0.05.

In addition to the analysis of power for each of these ten tests, confidence intervals for the linear contrast were computed from each set of simulated data. For the multivariate and mixed models approaches, these intervals were computed by using (3.8) and (3.11), respectively where the four estimates of degrees of freedom used previously determined the values of $t_{\alpha/2,\hat{\nu}}$ in the confidence interval calculations. The observed confidence levels and average confidence interval lengths were computed for each of the methods. Since confidence level and confidence interval length are each independent of the actual treatment difference values, the estimated confidence levels and average confidence interval lengths were computed by pooling information across all of the treatment differences and carryover differences. As a consequence these two characteristics are based on 13,000 simulations using model (1.1) and 36,000 simulations using model (1.2).

For crossover design models (1.1) and (1.2), the following information about confidence levels and length of confidence intervals was observed. All of the multivariate methods yielded confidence levels in excess of the nominal 0.95 level. Among the mixed models methods, T6 yielded a confidence level that was closest to 0.95. All four of the confidence levels for the mixed models approach were below 0.95 however.

Table 10 shows observed confidence levels for different sample sizes and different values of ε for the p=t=4 Williams' design. Table 11 show confidence levels (Z, TU, T2, T6) for the model with carryover parameters (1.2) for both treatment and carryover differences. An examination of Tables 10 and 11 reveal that T3 and T2, respectively, produce intervals that have observed confidence levels closest to 0.95, and consequently, each is the recommended method for the model indicated in the tables.

Tables 12 and 13 show the observed interval lengths for the models without carryover and with carryover, respectively. The observed lengths based on Z are also included for comparison purposes.

4.6 Example

Consider again the dairy science example. First, type H structure of Σ was tested.). It was found the Σ does not have type H structure, ($\chi_2^2 = 15.05$, p=0.0005, $\hat{\varepsilon}_2 = 0.8542$). Tests of equal treatment, equal period, and equal carryover effects were each conducted using the usual ANOVA F-test, the Huynh-Feldt test, a multivariate approach, and a mixed models approach. Additionally, pairwise comparisons of treatment effects and carryover effects were tested. All

test procedures were performed using a SAS program by Goad and Johnson (1993). The test statistics and their significance levels appear in Table 14.

5. Conclusions

Since a crossover design is a special form of a repeated measures design, then one should take the variance-covariance structure of the repeated measures into consideration when analyzing the collected data. Three alternative approaches to the analysis of a crossover experiment were proposed for cases when the variance-covariance matrix does not have type H structure.

If Σ does not have type H structure, of the methods presented here, the following recommendations are made. For a model with no carryover parameters, a recommended test for equal treatment effects is to use the Huynh-Feldt adjusted degrees of freedom method. For tests or confidence interval estimates of a linear contrast of the treatment parameters, the recommendation based on the simulation study is to use the multivariate method T3. For a model that does contain carryover parameters, multivariate method F1 is recommended for the test of equal treatment effects. The usual ANOVA F-test for equal carryover effects is recommended. For the test of a linear contrast of the treatment parameters, a multivariate approach (T2) is recommended. For the test of a linear contrast of the carryover parameters, the usual t-test is recommended. In all cases, the mixed models approach exhibits large test sizes, and the authors are reluctant to recommend this approach.

References

Box, G. E. P. 'Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification', *Annals of Mathematical Statistics*, **25**, 290-302, 1954.

Box, G. E. P. 'Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and correlation between errors in the two-way classification', *Annals of Mathematical Statistics* **25**, 484-498, 1954.

Cochran, William G. and Cox, Gertrude M. *Experimental Designs*, John Wiley & Sons: New York, 1957.

Graybill, F. A. *Introduction to Matrices with Applications in Statistics*. Wadsworth Publishing Co. Inc.: Belmont, California, 1969.

Graybill, F. A. *Theory and Application of the Linear Model*. Wadsworth Publishing Co. Inc.: Belmont, California, 1976.

Geisser, S. and Greenhouse, S. 'Extension of Box's results on the use of the F distribution in multivariate analysis', *Annals of Mathematical Statistics*, **29**, 885-891, 1958.

Greenhouse, S. W. and Geisser, S. 'On methods in the analysis of profile data', *Psychometrika*, **24**, 95-112, 1959.

Goad, C. L. 'Multivariate and mixed model analyses of crossover designs with more than two periods', Ph.D. Dissertation, Kansas State University, 1994.

Goad, C. L. and Johnson, D. E. 'On Multivariate Analyses of Crossover Designs', *Proceedings of the 1993 Kansas State University Conference on Applied Statistics in Agriculture*, 1993.

Goad, C. L. and Johnson, D. E. 'The Validity of ANOVA F-tests in Crossover Designs', submitted for publication.

Huynh, H. and Feldt, L. S. 'Conditions under which mean square ratios in repeated measurements designs have exact F-distributions', *Journal of the American Statistical Association*, **65**, 1582-1589, 1970.

Huynh, H. and Feldt, L. S. 'Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs', *Journal of Education Statistics*, **1**, 69-82, 1976.

Mathai, A. M. 'Moments of the trace of a non-central Wishart matrix' *Communications in Statistics, Theory and Methods*, **A9 (8)**, 795-801, 1980.

Jones, Byron, and Kenward, Michael G. *Design and Analysis of Cross-Over Trials.* Chapman and Hall: New York, 1989.

Ratkowsky, David A., Evans, Marc A., and Alldredge, J. Richard. *Cross-over Experiments, Design, Analysis, and Application.* Marcel Dekker, Inc.: New York, 1993.

Satterthwaite, F. E. 'An approximate distribution of estimates of variance components' *Biometrics Bulletin*, **2**:110-114, 1946.

Williams, E. J. 'Experimental designs balanced for the estimation of residual effects of treatments', *Australian Journal of Scientific Research*, **2**, 149-164, 1949.

Table 1: Dairy Science Example (p=t=3 Williams' Design)

	Diet	Period	Cow 1	2	3	4	5	6	7
Sequence 1	A	1	60.7	59.7	61.9	61.8	62.6		
	B	2	52.5	49.8	55.3	59.5	60.1		
	C	3	47.6	44.4	50.7	55.2	54.6		
Sequence 2	A	1	62.5	57.9	59.0	59.6	57.1	57.2	
	C	2	58.6	57.1	55.2	62.3	57.5	58.2	
	B	3	47.3	49.4	47.4	54.1	48.8	49.2	
Sequence 3	B	1	55.6	52.9	52.7	53.1	60.5	58.0	
	A	2	58.6	51.9	51.2	46.8	60.4	50.7	
	C	3	57.3	50.5	47.8	44.5	59.8	48.4	
Sequence 4	B	1	60.5	53.3	59.2	56.2	50.5	56.0	54.9
	C	2	59.8	53.6	57.8	53.0	55.4	54.4	54.1
	A	3	52.3	47.6	57.3	46.2	54.5	48.1	44.7
Sequence 5	C	1	62.8	55.8	56.0	62.3	60.1		
	A	2	55.6	52.1	48.2	55.1	54.6		
	B	3	57.8	59.1	52.4	62.8	60.5		
Sequence 6	C	1	54.9	60.6	55.9				
	B	2	45.7	52.6	46.0				
	A	3	52.8	50.3	50.6				

Table 2: 3 x 3 Williams' Design (MOLS)

Period	Sequence 1	2	3	4	5	6
1	A	A	B	B	C	C
2	B	C	A	C	A	B
3	C	B	C	A	B	A

Table 3: Four Treatment / Four Period Crossover Design

Period	Sequences			
	1	2	3	4
1	A	B	C	D
2	B	C	D	A
3	D	A	B	C
4	C	D	A	B

Table 4: Four Treatment / Four Period Design:
Σ Matrices Used in the Simulation,
$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$

$\rho_{12} = \rho_{23} = \rho_{34}$	$\rho_{13} = \rho_{24}$	ρ_{14}	ε
0.8	0.4	0.05	0.4799
0.8	0.6	0.10	0.4727
0.8	0.6	0.20	0.5238
0.8	0.6	0.40	0.6410
0.8	0.7	0.60	0.8205
0.8	0.8	0.80	1.0000

Table 5: Three Treatment / Three Period Design:
Σ Matrices Used in the Simulation,
$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$ and $\rho_{12} = \rho_{23} = 0.75$

ρ_{13}	ε
0.15	0.5586
0.25	0.6098
0.25	0.6739
0.50	0.8000
0.75	1.0000

Table 6: 3 x 3 Williams' Design Simulation Conditions

Variance-Covariance Matrix ε	No Carryover Model	Carryover Model		
0.5586	333333 555555 777777 999999	333333 233333 223333 233233	555555 455555 445555 355555 335555	999333 939339*
0.6098	333333 555555 777777 999999			
0.6739	333333 555555 777777 999999			
0.8000	333333 555555 777777 999999			
1.0000	333333 555555 777777 999999	333333 233233	555555	999333 939339*

*one Latin square with n_i=9 and second Latin square with n_i=3

Table 7: 4 x 4 Williams' Design Simulation
Conditions

Variance-Covariance Matrix ε	No Carryover Model
0.4799	3333
	5555
	7777
	9999
0.4727	3333
	5555
	7777
	9999
0.5238	3333
	5555
	7777
	9999
0.6410	3333
	5555
	7777
	9999
0.8205	3333
	5555
	7777
	9999
1.0000	3333
	5555
	7777
	9999

Table 8: Notation Used to Describe the Tests of Equal Treatment and Carryover Effects

Test	Description
F	Usual analysis of variance test
Adjustment to the Degrees of Freedom	
BOX	Conservative estimate for Box's ε ($\varepsilon = 1/(t-1)$)
GG	Greenhouse-Geisser adjustment, $\hat{\varepsilon}_1$
HF	Huynh-Feldt adjustment, $\hat{\varepsilon}_2$
Multivariate Approach (df = denominator degrees of freedom)	
F1	$df = \min_i\{\hat{v}_i\}$
F2*	$df = \max_i\{\hat{v}_i\}$
F3	$df = $ lower bound for $\hat{v}_i$, N - s
F4	$df = 2(N - s)$
F5*	$df = 3(N - s)$
F6	$df = df$ due to error from the usual ANOVA
Mixed Models Approach	
F7	$df = \min_i\{\hat{v}_i\}$
F8*	$df = \max_i\{\hat{v}_i\}$
F9	$df = $ lower bound for $\hat{v}_i$, N - s
F10	$df = 2(N - s)$
F11*	$df = 3(N - s)$
F12	$df = df$ due to error from the usual ANOVA

* *not applicable to p=t=3 Williams' design*

Table 9: Notation Used to Describe the Tests of Linear Contrasts,
Confidence Intervals, and Confidence Interval Lengths

Test	Description
Z	Standard normal test method, actual value of Σ used
TU	Usual t-test
Multivariate Approach (df = degrees of freedom for approximate t-tests)	
T1	$df = \hat{v}$
T2	$df = N - s$, lower bound for $\hat{v}$
T3	$df = 2(N - s)$, upper bound for $\hat{v}$
T4	$df = df$ due to error from the usual ANOVA
Mixed Models Approach	
T5	$df = \hat{v}$
T6	$df = N - s$, lower bound for $\hat{v}$
T7	$df = 2(N - s)$, upper bound for $\hat{v}$
T8	$df = df$ due to error from the usual ANOVA

Table 10: Confidence Levels of Preferred Methods -
Model with No Carryover Parameters
p=t=4 Williams' Design

ε	Sample	Z	TU	T3	T6
0.4727	3333	0.9488	0.9780	0.9523	0.9097
	5555	0.9492	0.9764	0.9545	0.9345
	7777	0.9455	0.9732	0.9479	0.9372
	9999	0.9508	0.9742	0.9523	0.9421
0.4799	3333	0.9505	0.9737	0.9593	0.9278
	5555	0.9488	0.9677	0.9546	0.9398
	7777	0.9510	0.9676	0.9551	0.9416
	9999	0.9505	0.9682	0.9512	0.9444
0.5238	3333	0.9479	0.9752	0.9592	0.9253
	5555	0.9503	0.9742	0.9572	0.9358
	7777	0.9494	0.9723	0.9512	0.9394
	9999	0.9488	0.9718	0.9502	0.9445
0.6410	3333	0.9508	0.9735	0.9662	0.9252
	5555	0.9487	0.9703	0.9602	0.9412
	7777	0.9501	0.9659	0.9563	0.9412
	9999	0.9490	0.9658	0.9552	0.9443
0.8205	3333	0.9482	0.9718	0.9688	0.9275
	5555	0.9518	0.9663	0.9616	0.9425
	7777	0.9503	0.9667	0.9589	0.9418
	9999	0.9506	0.9633	0.9575	0.9448
1.0000	3333	0.9505	0.9668	0.9688	0.9263
	5555	0.9532	0.9619	0.9622	0.9416
	7777	0.9502	0.9583	0.9587	0.9420
	9999	0.9523	0.9579	0.9577	0.9445

Table 11: Confidence Level of Preferred Methods - Model with Carryover Parameters
p = t = 3 Williams' Design

ε	Sample	Treatment				Carryover			
		Z	TU	T2	T6	Z	TU	T2	T6
0.5586	223333	0.9509	0.9093	0.9578	0.9338	0.9505	0.9586	0.9573	0.9344
	233233	0.9491	0.9115	0.9584	0.9337	0.9488	0.9566	0.9569	0.9344
	233333	0.9492	0.9126	0.9567	0.9338	0.9487	0.9541	0.9551	0.9351
	333333	0.9527	0.9166	0.9577	0.9371	0.9511	0.9565	0.9574	0.9351
	335555	0.9516	0.9063	0.9549	0.9423	0.9508	0.9544	0.9528	0.9444
	355555	0.9506	0.9106	0.9535	0.9412	0.9507	0.9504	0.9527	0.9439
	445555	0.9499	0.9076	0.9536	0.9406	0.9485	0.9521	0.9518	0.9403
	455555	0.9524	0.9139	0.9558	0.9419	0.9496	0.9529	0.9531	0.9415
	555555	0.9509	0.9128	0.9553	0.9438	0.9501	0.9516	0.9516	0.9416
	939339	0.9503	0.9109	0.9516	0.9441	0.9511	0.9669	0.9523	0.9441
	999333	0.9518	0.9324	0.9528	0.9437	0.9490	0.9524	0.9699	0.9438

Table 12: Interval Lengths of Preferred Methods -
Model with No Carryover Parameters
p=t=4 Williams' Design

ε	Sample	Z	TU	T3	T6
0.4727	3333	0.8765	1.0618	0.9565	0.6810
	5555	0.6790	0.7997	0.7111	0.5040
	7777	0.5738	0.6649	0.5897	0.4191
	9999	0.5061	0.5845	0.5181	0.3678
0.4799	3333	1.0121	1.1687	1.1192	0.6706
	5555	0.7840	0.8767	0.8257	0.4894
	7777	0.6626	0.7309	0.6858	0.4048
	9999	0.5843	0.6392	0.5991	0.6392
0.5238	3333	0.8765	1.0556	0.9790	0.7376
	5555	0.6790	0.7866	0.7175	0.5381
	7777	0.5738	0.6555	0.5955	0.4465
	9999	0.5061	0.5738	0.5205	0.3909
0.6410	3333	0.8765	1.0264	1.0017	0.8077
	5555	0.6790	0.7612	0.7289	0.588.
	7777	0.5738	0.6308	0.6012	0.4877
	9999	0.5061	0.5504	0.5237	0.4252
0.8205	3333	0.8002	0.9295	0.9265	0.8383
	5555	0.6198	0.6839	0.6679	0.6093
	7777	0.5238	0.5677	0.5519	0.5063
	9999	0.4620	0.4941	0.4796	0.4414
1.0000	3333	0.7157	0.8082	0.8292	0.8029
	5555	0.5544	0.5938	0.5971	0.5844
	7777	0.4685	0.4918	0.4932	0.4857
	9999	0.4132	0.4288	0.4295	0.4244

Table 13:   Interval Lengths of Preferred Methods - Model with Carryover Parameters
p = t = 3 Williams' Design

		Treatment				Carryover			
ε	Sample	Z	TU	T2	T6	Z	TU	T2	T6
0.5586	223333	1.2706	1.1318	1.4515	0.6599	1.4466	1.5396	1.6415	0.8005
	233233	1.2706	1.1340	1.4549	0.6604	1.4466	1.5425	1.6460	0.7996
	233333	1.1934	1.0729	1.3487	0.6285	1.4187	1.4832	1.5940	0.7406
	333333	1.1410	1.0270	1.2745	0.6193	1.3148	1.3778	1.4615	0.7162
	335555	1.0155	0.8802	1.0844	0.4985	1.1526	1.2016	1.2269	0.6130
	355555	0.9375	0.8221	0.9951	0.4718	1.1245	1.1465	1.1899	0.5577
	445555	0.9353	0.8179	0.9922	0.4792	1.0707	1.1056	1.1325	0.5669
	455555	0.9043	0.7956	0.9572	0.4679	1.0594	1.0844	1.1183	0.5452
	555555	0.8838	0.7795	0.9335	0.4648	1.0184	1.0458	0.8875	0.5363
	939339	0.9791	0.8555	1.0213	0.5018	0.9383	1.0355	0.9521	0.5639
	999333	0.8497	0.8008	0.8875	0.5037	0.9297	0.9546	0.9699	0.5271

Table 14: Test Statistics and Their Significance Levels for the Dairy Science Example

Effects	Usual ANOVA	Huynh-Feldt	Multivariate, F1	Mixed Models, F9
		Methods		
Diet	4.00	4.00	4.3865*	3.1362
	p=0.0236	p=0.0511	p=0.0228	p=0.0602
Period	47.97	47.97	288.2389	290.5940
	p<0.0001	p<0.0001	p<0.0001	p<0.0001
Carryover	22.40*	22.40	32.8925	55.7624
	p<0.0001	p<0.0001	p<0.0001	p<0.0001

Effect	Usual t	Multivariate, T2	Mixed Models, T6
	Methods for Linear Contrasts		
Diet			
A - B	2.67	2.80*	2.46
	p=0.0098	p=0.0095	p=0.0105
A - C	2.14	2.06*	1.64
	p=0.0364	p=0.0493	p=0.1124
B - C	-0.38	-0.41*	-0.83
	p=0.7084	p=0.6857	p=0.4147
Carryover			
A - B	2.98*	3.83	3.06
	p=0.0042	p=0.0008	p=0.0050
A - C	6.69*	8.33	10.18
	p<0.0001	p<0.0001	p<0.0001
B - C	3.86*	4.33	7.84
	p=0.0003	p=0.0002	p<0.0001

* recommended test procedures for a model containing carryover parameters

Figure 1:   Power Curves for Tests of Equal Treatment Effects

Model with No Carryover Parameters
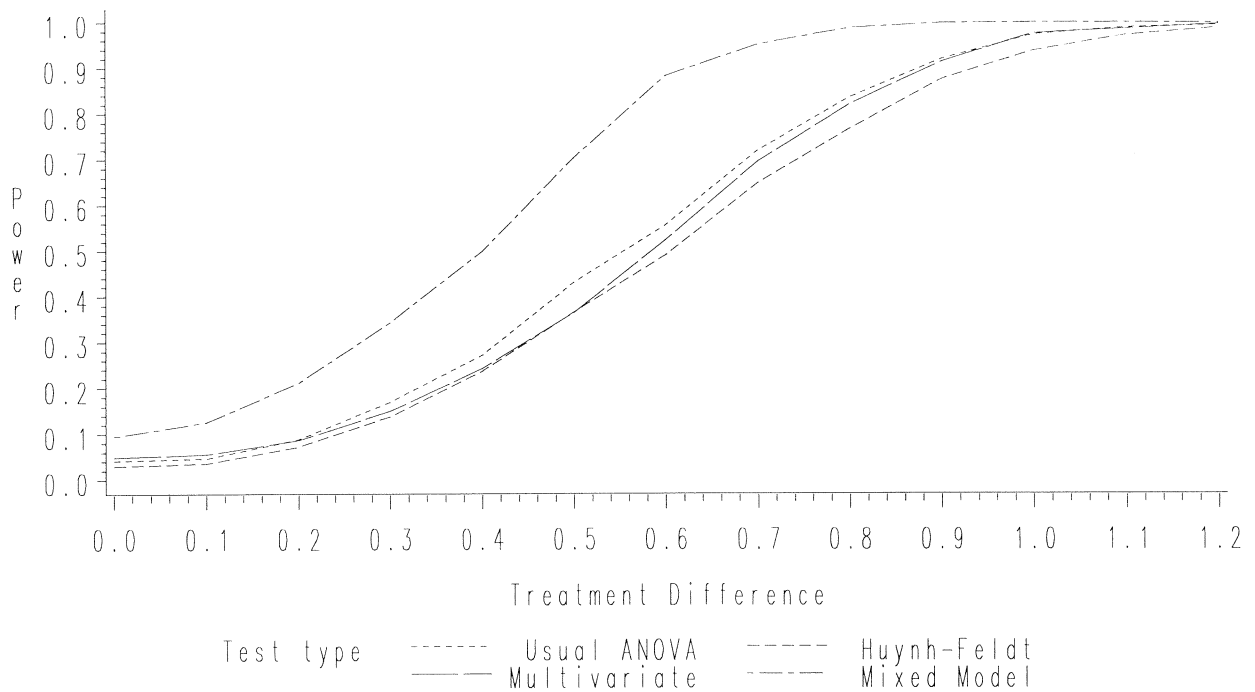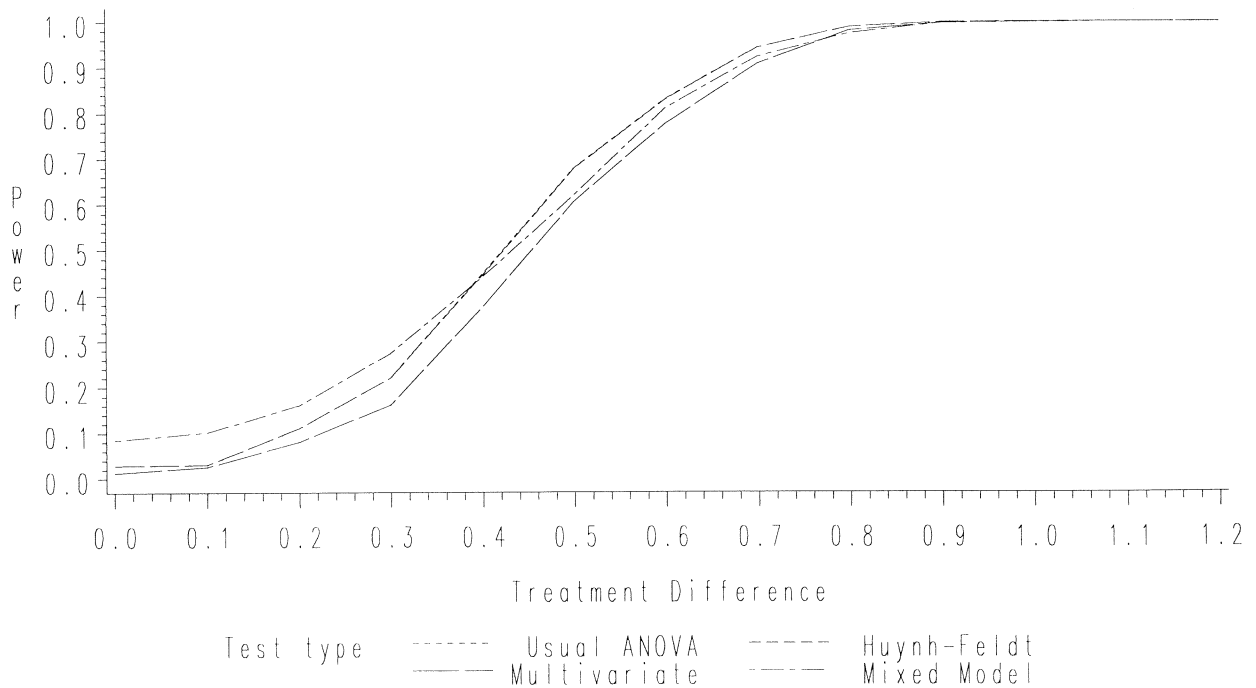$\varepsilon = 0.4727$ Sample = 3333

Treatment Difference

Test type   ------- Usual ANOVA   ----- Huynh-Feldt
——— Multivariate   -----— Mixed Model

Figure 2:   Power Curves for Tests of Equal Treatment Effects

Model with No Carryover Parameters
$\varepsilon = 1.0000$ Sample = 3333

Treatment Difference

Test type   ------- Usual ANOVA   ----- Huynh-Feldt
——— Multivariate   -----— Mixed Model

Figure 3: Power Curves for Tests of Equal Treatment Effects

Model with Carryover Parameters
ε = 0.5586   Sample = 333333

Treatment Difference

Test type   ------ Usual ANOVA   ---- Huynh-Feldt   ------ Multivariate   ----- Mixed Model
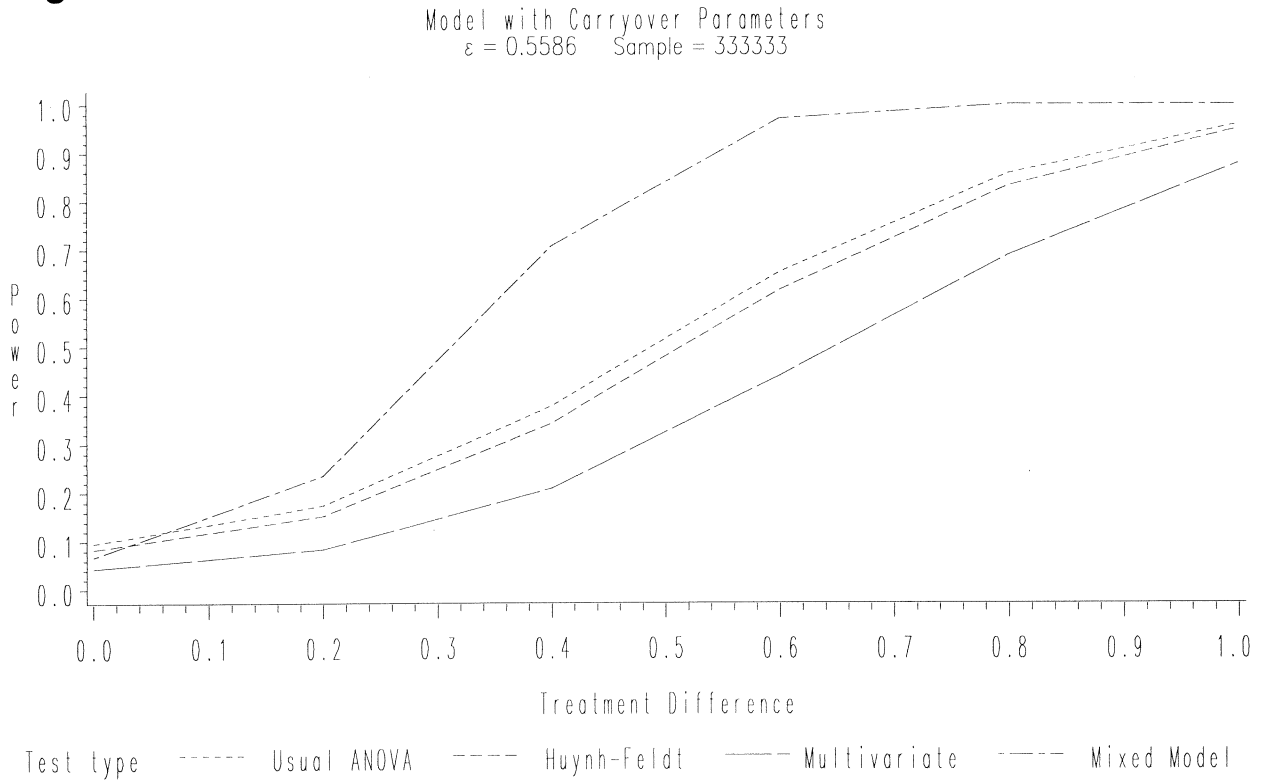
Figure 4: Power Curves for Tests of Equal Carryover Effects

Model with Carryover Parameters
ε = 0.5586   Sample = 333333

Carryover Difference

Test type   ------ Usual ANOVA   ---- Huynh-Feldt   ------ Multivariate   ----- Mixed Model
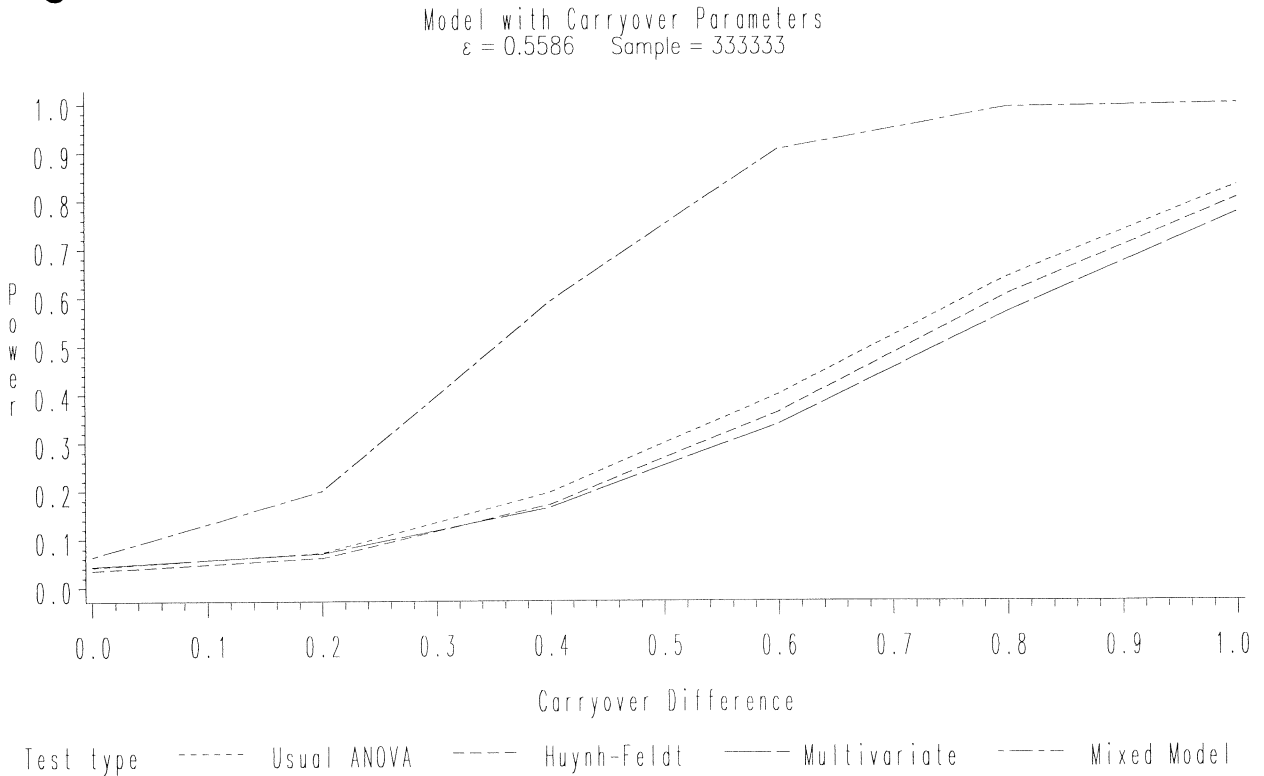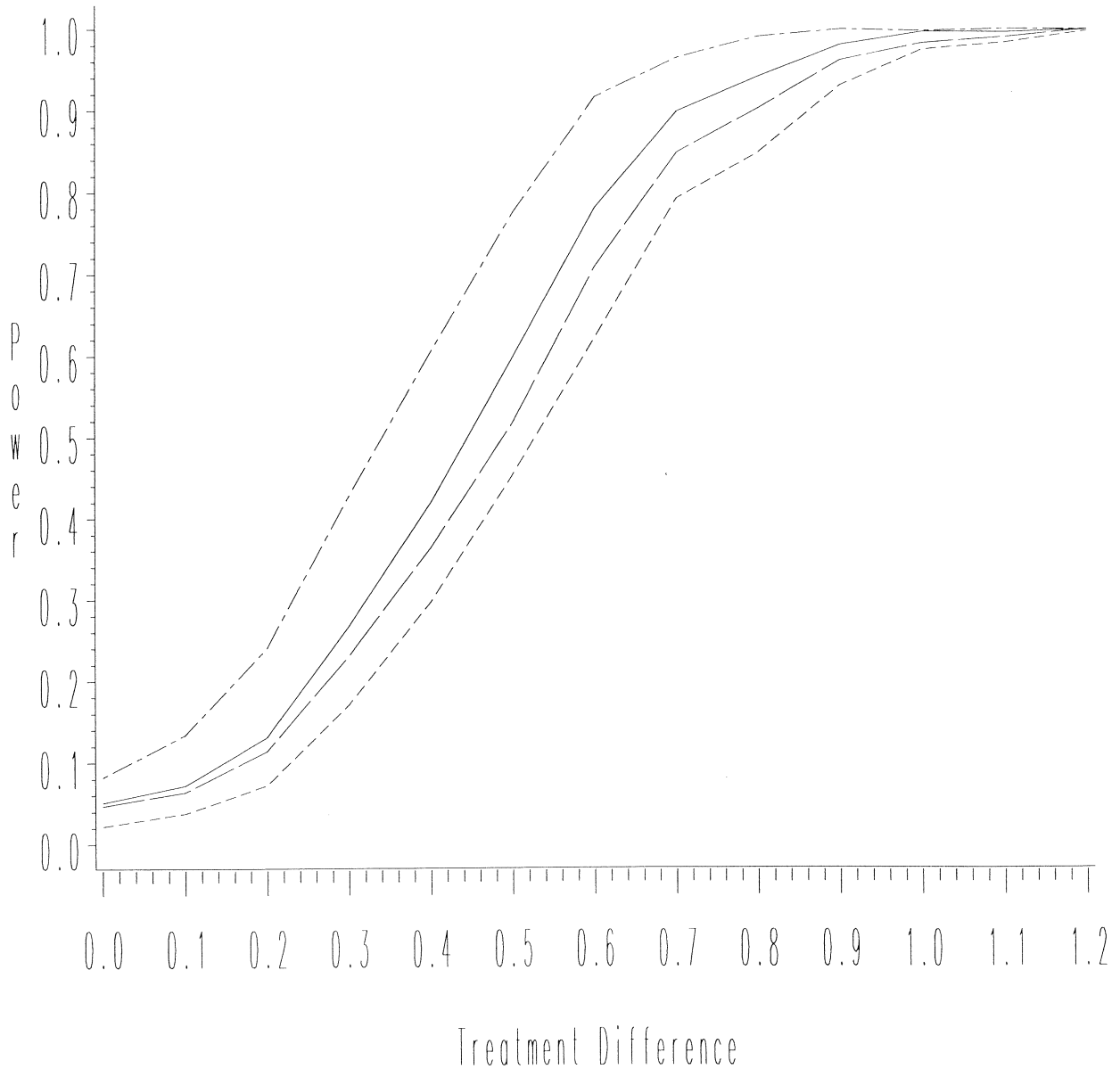
Figure 5: Power Curves for a Linear Contrast of Treatment Effects

Model with No Carryover Parameters
$\varepsilon = 0.4727$ Sample = 3333