STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:2585–2607 (DOI: 10.1002/sim.731)

Non-parametric analysis of covariance for confirmatory randomized clinical trials to evaluate dose–response relationships

Catherine M. Tangen[1,*,†] and Gary G. Koch[2,3]

[1]*Southwest Oncology Group, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., MP-557, P.O. Box 19024, Seattle, WA 98109-1024, U.S.A.*
[2]*Department of Biostatistics, CB 7400, School of Public Health, University of North Carolina, Chapel Hill, NC 27599-7400, U.S.A.*
[3]*DeMontfort University, Faculty of Computing Sciences and Engineering, Department of Medical Statistics, The Gateway, Leicester, LE1 9BH, U.K.*

SUMMARY

In confirmatory randomized clinical trials that are designed to compare multiple doses of a test treatment with a control group and with one another, there are often statistical issues regarding compound hypotheses and multiple comparisons which need to be considered. In most cases the analysis plan needs a clear specification for the proposed order for conducting statistical tests (or for managing the overall significance level), which statistical methods will be used, and whether adjustment for covariates will be performed. There are several benefits of specifying non-parametric analysis of covariance (ANCOVA) for performing the primary confirmatory analyses. Only minimal assumptions are needed beyond randomization in the study design, whereas regression model based methods have assumptions about model fit for which departures may require modifications that are incompatible with a fully prespecified analysis plan. Non-parametric methods provide traditionally expected results of ANCOVA; namely, a typically small adjustment to the estimate for a treatment comparison (so as to account for random imbalance of covariates between treatment groups) and variance reduction for this estimate when covariates are strongly correlated with the response of interest. The application of non-parametric ANCOVA is illustrated for two randomized clinical trials. The first has a (3×4) factorial response surface design for the comparison of 12 treatments (that is, combinations of three doses of one drug and four doses of a second drug) for change in blood pressure; and the second example addresses the comparison of three doses of test treatment and placebo for time-to-disease progression. This clinical trial has comparisons among treatments made for a dichotomous criterion, Wilcoxon rank scores and averages of cumulative survival rates. In each example, the non-parametric covariance method provides variance reduction relative to its unadjusted counterpart. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

This paper discusses the role of non-parametric analysis of covariance in confirmatory clinical trials to compare multiple doses of a test treatment with a control group and with one another.

*Correspondence to: Catherine M. Tangen, Southwest Oncology Group, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., MP-557, P.O. Box 19024, Seattle, WA 98109-1024, U.S.A.
†E-mail: cathyt@swog.fhcrc.org

For such confirmatory clinical trials, one is usually interested in whether one or more doses is better than the control and whether higher doses are better than lower doses. For this purpose, compound hypotheses with possibly multiple comparisons require evaluation, and a key issue is the control of the experimentwise significance level. Non-parametric analysis of covariance can strengthen the analysis of these types of clinical trials by providing more precise estimates for linear contrasts among treatments (such as high dose–placebo or the average of high and medium dose–placebo) than an unadjusted analysis, particularly when covariables are strongly correlated with response variables. It also can clarify interpretation by providing estimates for treatment comparisons with removal of the influence of random imbalances for the distributions of covariables. Additional advantages are its use of essentially no assumptions in studies with randomized assignment of treatments and the similar nature of its applicability to continuous measurements, ranks for ordinal outcomes, or dichotomous outcomes for either a univariate response variable or a multivariate profile of response variables [1–8].

The most straightforward form of non-parametric analysis of covariance focuses on pairwise differences between each dose and control for vectors of means of response variables and covariables, with any stratification of randomization in the design of the clinical trial being either ignored or incorporated in the covariables. A linear model which specifies 0's for differences between treatments for means of covariables is fit to this set of vectors by weighted least squares. The resulting parameter estimates are covariance adjusted estimates for differences in means for response variables, and they have the advantage of reduced variance relative to their unadjusted counterparts [2]. Through these adjusted means, comparisons among treatment groups can have more powerful statistical tests or narrower confidence intervals. For confirmatory clinical trials with multiple doses of test treatment, variance reduction is important because statistical analysis requires a method to account for multiple comparisons, and because comparisons of lower doses with higher doses usually involve smaller true differences than comparisons of higher doses with control. When adjustment for stratification in the design of such clinical trials is necessary, extensions of the methods summarized here are available through their application to weighted combinations of vectors of means across strata (see reference [2] and Appendix A1).

The application of non-parametric analysis of covariance to comparisons among multiple doses of a test treatment and placebo is illustrated for two confirmatory randomized clinical trials. The first example has a (3×4) factorial response surface design for the comparison of 12 treatments (that is, placebo, two doses of hydrochlorothiazide (HCTZ), three doses of an angiotensin converting enzyme inhibitor (ACEI), and six combinations of HCTZ and ACEI) for change in resting systolic and diastolic blood pressure relative to baseline [9]. The second example is a confirmatory clinical trial to compare three doses of test treatment and placebo for time-to-disease progression. Comparisons among treatments are illustrated for a dichotomous criterion, Wilcoxon rank scores and logrank scores, and averages of cumulative survival rates. For each of these examples, non-parametric covariance adjustment provided substantial variance reduction and narrower confidence intervals than its unadjusted counterpart.

Extensive background on the underlying principles for non-parametric ANCOVA behind this method for comparisons between two groups for means of response variables can be found in Koch *et al.* [2]. In that paper, there is additional discussion about the advantages and limitations of these methods in contrast to more traditional regression model-based methods as well as the complementary and mutually reinforcing roles that both types of methods can serve. In some clinical trials, the criteria of interest can be functions of means such as

odds ratios for dichotomous and ordinal outcomes as well as incidence density ratios and cumulative survival rates for times-to-events. Other references describe the non-parametric covariance method for these types of criteria in the case of two treatment groups [4–6, 8].

2. DESIGN AND ANALYSIS ISSUES

There are several different designs that can be implemented in multi-dose clinical trials. Most designs belong to the following categories: (i) two or more doses of test drug and placebo; (ii) two or more doses of test drug and one or more doses of active reference control drug; (iii) three or more doses of test drug (with no control drug); (iv) two or more doses of test drug, placebo, and one or more doses of active reference control drug, or (v) a response surface design for combinations of two or more doses of test drug with one or more doses of active reference control drug, their corresponding monotherapies, and placebo [7, 9–14].

A set of specific statistical issues for multi-dose clinical trials involve multiplicity for treatment comparisons. Multiple comparisons between doses of test drug and control are of the most interest with the 'control' being a placebo group, doses of active reference control or lower doses of test drug. Multiplicity also occurs when there are multiple comparisons between higher and lower doses of the test drug, and there can be multiple comparisons to evaluate non-inferiority of better doses of test drug to doses of active reference control [15–25].

There are at least six inferential postures for multi-dose clinical trials. The first posture is mainly exploratory. There are no real underlying assumptions for this posture, except that low doses are considered safe and tolerable. The sample size for this posture is usually small and has flexibility for increases in a *post hoc* manner. This posture is typical for a phase IIA clinical trial. The statistical section of the analysis plan or protocol can be minimal for this posture, but guidelines can be helpful to support appropriately cautious interpretation. Conclusions based on this posture can be descriptively straightforward, but caution is required when there is no planned method for managing multiplicity.

Posture 2 is confirmatory for global assessment of favourable activity of test drug through a prespecified contrast involving a trend for two or more doses versus control, but it is usually exploratory for all other comparisons among treatments. Underlying assumptions for this posture include all doses under consideration being safe and tolerable and the expectation of more favourable responses at higher doses (that is, monotonicity for efficacy). The analysis plan must clearly state the method of evaluation for a prespecified contrast for trend. Moderate sample sizes are required to have sufficient power for the prespecified contrast, and conclusions are usually inferentially confirmatory and straightforward for prespecified contrasts, but descriptive and cautious otherwise. This posture is useful for a phase IIB clinical trial.

Posture 3 is similar to posture 2 in that it is confirmatory for trend as in posture 2, but it is also confirmatory for pairwise comparisons between test drug and placebo; for all other comparisons among treatments, it is exploratory. The underlying assumptions for this third posture are that at least one of the higher doses is efficacious, safe and tolerable, but some lower doses may not have sufficient efficacy, and the highest dose may not be sufficiently tolerable to have adequate compliance for demonstrating efficacy. Sample sizes need to be large in order to have sufficient power for comparisons between one or more separate doses and placebo (phase 3) with a method which clearly manages multiplicity through more stringent significance levels than 0.05. The analysis plan needs to state clearly the methods for

managing both a global assessment of activity and comparisons between the separate doses of test drug and placebo. The conclusions are inferentially confirmatory and straightforward for the global assessment of activity and the comparisons of separate doses of test drug to placebo, but descriptive and cautious otherwise.

Posture 4 has the same objectives as posture 3 and additionally is confirmatory for pairwise comparisons to demonstrate non-inferiority of doses of test drug to doses of active reference control drug, but it is exploratory for all other comparisons. Posture 5 is confirmatory for the objectives in posture 3 and for prespecified contrasts to demonstrate more favourable activity for some doses of test drug than others, but exploratory for all other comparisons. Posture 6 is confirmatory for the objectives of both posture 4 and posture 5. Postures 4–6 often arise for phase IV clinical trials where the relationship between doses of test drug and response criteria are considered well understood from prior studies for test drug or other drugs in the same class (that is, there is reasonably sound prior knowledge of the doses which have better efficacy as well as sufficient safety and tolerability). Very large sample sizes are needed to have sufficient power for all comparisons. The analysis section of the protocol should have extensive details for the methods for all comparisons and substantial discussion of the management of multiplicity. The results for this type of study can be inferentially confirmatory, but possibly complex, for the comparisons which are well managed by its methods to address multiplicity; they are typically straightforward for the types of inferential assessments that are shared with posture 3.

Difficulties can be encountered for the design and analysis plans for pairwise comparisons between doses for postures 3–6. There is reduced power from smaller sample sizes, reduced power from possibly more stringent significance levels to account for multiplicity, and reduced power from smaller true differences in comparisons of doses with one another and in comparisons of some doses to placebo. Senn *et al*. provide a contrasting perspective on postures for dose-finding trials [14].

2.1. Analysis plans

The first priority for analysis plans should be to have inferential confirmation for significantly favourable activity in the relationship between doses of test drug and response through a prespecified contrast involving two or more doses versus control (trend). The second priority is inferential confirmation that at least one dose of test drug is superior to placebo. In some studies, a third priority is inferential confirmation that at least one of the doses of test drug which is superior to placebo is also non-inferior to active reference control. In others, there is interest for whether one of the doses which is superior to placebo fulfils other criteria which other doses do not. For example, there could be a requirement for the point estimate for effect size to exceed a prespecified threshold for clinical relevance [7].

A strategic order can be specified for the comparisons in the analysis plan. It would usually have a global assessment first; then given significance for this assessment, comparisons to placebo would then have evaluation; other comparisons would have assessment at the third level. Methods for managing multiplicity within each of the respective steps of this prespecified order are necessary [15–25]. Since the significance levels from methods to address multiplicity can be more stringent than 0.05, one should apply strategies to make comparisons as powerful as possible. These include covariance adjustment, composite endpoints, and planned meta-analyses for two or more studies with the same protocol [10, 20].

Several methods are available for global assessments. One possibility is a prespecified contrast involving two or more doses of test drug versus control. Their scope includes: a linear contrast for better response with increasing dose (for example, $3 \times$ high + middle − low − $3 \times$ placebo); an average of all doses versus placebo (for example, high + middle + low − $3 \times$ placebo); or the average of the two highest doses versus placebo (for example, high + middle − $2 \times$ placebo). Another option is to have the comparison between the highest dose and placebo (for example, high − placebo) be the global assessment.

There are several strategies for managing multiplicity within a particular step of the analysis. One method is to prespecify two contrasts and test these contrasts with a bivariate test statistic with two degrees of freedom (for example, high dose − placebo, middle dose − placebo). If it has significance at the 0.05 level, then each of the separate contrasts can have assessment at the 0.05 level [15]. Another strategy is to implement more powerful versions of the Bonferroni method such as the Hochberg method [16, 22]; for example, assess the largest p-value for c tests at α; if non-significant, assess the next largest at $(\alpha/2)$ etc. A third strategy is to prespecify the order for comparisons and to proceed with statistical tests in that order as long as all previous tests are significant; for example, (high − placebo) first; and if significant, (middle − placebo) etc. [15, 20].

For whatever method is used to manage multiplicity, re-sampling procedures for enumeration of simulations of the randomization distribution can be useful for determining approximately exact p-values [24, 25].

2.2. *Roles for covariance adjustment*

Another strategy for increasing power for treatment comparisons is to implement analysis of covariance (ANCOVA). This method can provide more powerful statistical tests (or narrower confidence intervals) through 'variance reduction' in statistics for comparisons of randomized groups. It also provides a structure for which random imbalances for covariables are 'adjusted to equality'. In this way, analysis of covariance can clarify the degree to which detected differences between randomized groups are due to treatment rather than random imbalances for baseline factors which are associated with response. ANCOVA can also provide a structure for evaluating homogeneity of treatment differences across subgroups.

For ANCOVA, it is important to have *a priori* specification of covariates in order to avoid criticisms of potential bias from selection [26]. It is also important to specify covariates that are expected to have strong correlations with response criteria in order to obtain greater benefits of variance reduction and correspondingly increased power [2].

There are at least two different methods that can be employed to adjust for covariates. One frequently used option is a parametric method with a statistical regression model for the relationship between covariables and the conditional distributions of response variables given the covariables. Another option is a non-parametric method which involves a linear model for (unconditional) differences between treatment groups for response criteria and covariables jointly and with specifications that adjust differences between groups for means of covariables to zero. For hypothesis testing of no differences between treatment groups, the study design provides the basis for the distribution of results from the non-parametric method under the null hypothesis (see Section 3.3). When confidence intervals concerning treatment differences and tests concerning treatment $\times$ subgroup interaction are of interest, the principal assumption for the non-parametric method is that the patients for each treatment $\times$ subgroup are comparable to

a simple random sample from a corresponding conceptual population. The extent of potential interaction is often of interest for centres, demographic factors such as gender and age, and aspects of baseline severity of disease.

Parametric regression model fitting methods have several forms, and they include multiple linear regression for measurements on a continuous scale, multiple logistic regression for dichotomous response, proportional odds regression for ordinal response, and proportional hazards regression for data concerning times to event. However, there are certain dilemmas for regression model fitting methods. There can be uncertainty for the explanatory variables or interactions of these variables with treatment to include in the model and their specification (for example, age as linear, quadratic, or classes), and there can be uncertainty for the interactions among explanatory variables to include in the model. For ordinal or time to event response criteria, there are additional issues of concern including the appropriateness of assumptions [27], such as 'proportional odds' or 'proportional hazards', and the implications of inclusion of covariables in the model relative to the interpretation of the treatment parameter [28–30].

A way to address dilemmas for parametric regression models is the use of a non-parametric method for the primary confirmatory evaluation of treatment comparisons since it requires essentially no assumptions for a randomized clinical trial [1–8]. A complementary statistical regression model can then be used to describe the nature and extent of treatment effects and to evaluate the homogeneity of treatment effects across subgroups. With this strategy, *a posteriori* modification of a model is possible to improve its compatibility with the data, and more than one type of model can be assessed in order to provide confirmation for similar conclusions.

For non-parametric methods for primary confirmatory inference, it is necessary to prespecify both the method and the covariables for adjustment in the protocol (or in an amendment or formal plan prior to any unmasking of the study). In this way, the significance level is controlled, and estimation bias is avoided [31]. By adjusting for covariables which are expected to be strong correlates of response on the basis of previous studies, one can usually obtain a reduction in variance and gain power for the comparisons among treatment groups, and one can also account for the influence of random differences among groups for these covariables on the differences among groups for the response variables.

In supportive analyses, non-parametric methods can be used to evaluate robustness of conclusions to adjustment for other covariables. Some candidate covariables for these exploratory analyses would be those with atypical imbalances between the treatment groups, and covariables with strong association with response but which were not prespecified in the analysis plan. Since these analyses have a *post hoc* nature, one should view their results cautiously.

3. METHODS FOR NON-PARAMETRIC ANCOVA

Let $i = 0, 1, 2, \ldots, s$ index a set of $(s + 1)$ treatment groups to which patients are randomized (without any stratification) and let $k = 1, 2, \ldots, n_i$ index patients with treatment i; also, the 0th group corresponds to a control treatment such as placebo. Let $\boldsymbol{y}_{ik} = (y_{ik1}, \ldots, y_{ikr})'$ denote the vector of observed scores (or values) for r response variables for patient k with treatment i. The scores for the y_{ikj} could be measured values, ranks, or 0 or 1 indicators. From this notation we see that the methodology can be applied to either a univariate response or multivariate response. The matrix equations in the subsequent discussion function in the same way regardless of the number of responses. In our first example in Section 4.1, systolic and

diastolic blood pressure have a bivariate structure for the responses of interest, and for the second example in Section 4.2, survival rates for three intervals of time during the study have a trivariate structure.

Let $\boldsymbol{x}_{ik} = (x_{ik1}, \ldots, x_{ikm})'$ denote the vector of m pre-treatment (that is, background or baseline) covariables for patient k with treatment i. Let $\boldsymbol{f}_{ik} = (\boldsymbol{y}'_{ik}, \boldsymbol{x}'_{ik})'$ denote the concatenated vector of response variables and covariables for the kth patient with treatment i. The vectors $\boldsymbol{f}_{ik}$ are assumed to have no missing elements throughout this discussion; some ways to manage missing data for the methods described here are suggested in Koch *et al.* [2].

Let $\bar{\boldsymbol{f}}_i = \{\sum_{k=1}^{n_i} \boldsymbol{f}_{ik}/n_i\} = (\bar{\boldsymbol{y}}'_i, \bar{\boldsymbol{x}}'_i)'$ denote the concatenated vector of means of the response variables $\bar{\boldsymbol{y}}_i$ and covariables $\bar{\boldsymbol{x}}_i$ for the ith treatment, and let $\bar{\boldsymbol{f}} = (\bar{\boldsymbol{f}}'_0, \ldots, \bar{\boldsymbol{f}}'_s)'$ denote the concatenated vector of mean vectors for all $(s+1)$ treatments. A basis for comparisons among the $(s+1)$ groups is the set of differences $\boldsymbol{F}_i = (\bar{\boldsymbol{f}}_i - \bar{\boldsymbol{f}}_0)$ between the means for the ith group and the means for the 0th (or control) group for $i = 1, 2, \ldots, s$. The $\boldsymbol{F}_i$ have the general structure shown in (1)

$$\begin{aligned} \boldsymbol{F}_i = (\bar{\boldsymbol{f}}_i - \bar{\boldsymbol{f}}_0) &= ((\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}}_0)', (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_0)')' \\ &= (\boldsymbol{d}'_i, \boldsymbol{u}'_i)' \\ &= [-\boldsymbol{I}_{r+m}, \boldsymbol{I}_{r+m}][\bar{\boldsymbol{f}}'_0, \bar{\boldsymbol{f}}'_i]' \end{aligned} \tag{1}$$

where $\boldsymbol{I}_{r+m}$ is the $(r+m) \times (r+m)$ identity matrix.

Non-parametric covariance adjustment is applied to the $\boldsymbol{F}_i$ by fitting a linear model which specifies no differences among the groups for the means of the covariables. This linear model has the specification in (2)

$$\begin{aligned} \boldsymbol{F}_i = \begin{bmatrix} \boldsymbol{d}_i \\ \boldsymbol{u}_i \end{bmatrix} &\hat{=} \begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0}_{mr} \end{bmatrix} \boldsymbol{g}_i \quad \text{for } i = 1, 2, \ldots, s \\ &\hat{=} \boldsymbol{X}_1 \boldsymbol{g}_i \end{aligned} \tag{2}$$

where '$\hat{=}$' denotes 'is estimated by' and $\boldsymbol{0}_{mr}$ is an $(m \times r)$ matrix of 0s. In (2), the $\boldsymbol{g}_i$ represent adjusted differences between the ith group and the 0th group for the means of the response variables. The adjustment is through the equating of the $\boldsymbol{u}_i$ to $\boldsymbol{0}$'s by (2).

The corresponding specification in terms of the concatenated vector $\boldsymbol{F} = (\boldsymbol{F}'_1, \ldots, \boldsymbol{F}'_s)'$ is shown in (3):

$$\begin{aligned} \boldsymbol{F} = \begin{bmatrix} \bar{\boldsymbol{f}}_1 - \bar{\boldsymbol{f}}_0 \\ \cdots \\ \bar{\boldsymbol{f}}_s - \bar{\boldsymbol{f}}_0 \end{bmatrix} &\hat{=} \begin{bmatrix} \boldsymbol{X}_1 & \cdots & \boldsymbol{0}_{(r+m),r} \\ \cdots & \cdots & \cdots \\ \boldsymbol{0}_{(r+m),r} & \cdots & \boldsymbol{X}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{g}_1 \\ \cdots \\ \boldsymbol{g}_s \end{bmatrix} \\ \boldsymbol{F} = [\boldsymbol{L}_1 \otimes \boldsymbol{I}_{(r+m)}]\bar{\boldsymbol{f}} &\hat{=} [\boldsymbol{I}_s \otimes \boldsymbol{X}_1]\boldsymbol{g} \\ \boldsymbol{F} = \boldsymbol{L}\bar{\boldsymbol{f}} &\hat{=} \boldsymbol{X}\boldsymbol{g} \end{aligned} \tag{3}$$

with $\otimes$ denoting the Kronecker product matrix operation whereby the matrix on the right multiplies each of the elements of the matrix on the left, and $\boldsymbol{L}_1 = [-\boldsymbol{1}_s, \boldsymbol{I}_s]$.

Determination of $\boldsymbol{g}$ is through weighted least squares with weights based on the inverse of a consistent estimator for the covariance matrix $\boldsymbol{V_F}$ for $\boldsymbol{F}$, that is

$$\boldsymbol{g} = (\boldsymbol{X}'\boldsymbol{V_F}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V_F}^{-1}\boldsymbol{F} \tag{4}$$

Methods for obtaining $\boldsymbol{V_F}$ are discussed in Sections 3.2 and 3.3. A consistent estimator for the covariance matrix for $\boldsymbol{g}$ is $\boldsymbol{V_g} = (\boldsymbol{X}'\boldsymbol{V_F}^{-1}\boldsymbol{X})^{-1}$. Sufficiently large sample sizes $\{n_i\}$ enable $\boldsymbol{g}$ to have an approximately multivariate normal distribution for which the covariance matrix is essentially known through $\boldsymbol{V_g}$.

A test statistic for the linear hypothesis H_0:$\boldsymbol{Kg} \widehat{=} \boldsymbol{0}$, where $\boldsymbol{K}$ has full rank, is provided by

$$Q(\boldsymbol{Kg}) = \boldsymbol{g}'\boldsymbol{K}'(\boldsymbol{K}\boldsymbol{V_g}\boldsymbol{K}')^{-1}\boldsymbol{Kg} \tag{5}$$

Under H_0, this statistic approximately has the chi-squared distribution with degrees of freedom (d.f.) equal to rank $(\boldsymbol{K})$. For situations where sample sizes are not large enough to support such approximations, exact p-values for $Q(\boldsymbol{Kg})$ are available from its permutation distribution relative to all possible ways of assigning the $n = \sum_{i=0}^{s} n_i$ patients to the $(s+1)$ treatment groups; see Section 3.3 for further discussion of this point.

The extent to which covariance adjustment through (3) counteracts random imbalances among the $(s+1)$ groups for the distributions of the covariables is described by the goodness-of-fit statistic with $\hat{\boldsymbol{F}} = \boldsymbol{Xg}$

$$Q = (\boldsymbol{F} - \hat{\boldsymbol{F}})'\boldsymbol{V_F}^{-1}(\boldsymbol{F} - \hat{\boldsymbol{F}}) \tag{6}$$

where $\hat{\boldsymbol{F}} = \boldsymbol{Xg}$. This statistic approximately has the chi-squared distribution with d.f. $= sm$. Large values of Q, or correspondingly small p-values, express the extent of atypical random imbalances among the $(s+1)$ groups for the means of the m covariables. Conversely, more typically expected p-values (for example, $p > 0.10$) correspond to $\hat{\boldsymbol{F}}$ and $\boldsymbol{F}$ being reasonably similar, and so non-parametric covariance adjustment provides variance reduction through $\boldsymbol{V_g}$ without there being substantial differences between the adjusted estimates $\boldsymbol{g}$ for treatment differences and their unadjusted counterparts $\boldsymbol{d}$; see Appendix A3 for further clarification.

3.1. Restricted evaluation of specified contrasts

In situations with moderate, rather than clearly large, sample sizes (for example, $20 \leqslant n_i \leqslant 80$), there may be some limitations in the statistical behaviour of $\boldsymbol{g}$ from near (or actual) singularities in $\boldsymbol{V_F}$. This point merits particular attention when the dimension of the $\boldsymbol{F}_i$ (that is, $(r+m)$) exceeds $(0.5n_i)$ for some of the treatment groups. One way to address this difficulty is to apply non-parametric covariance adjustment in a separate manner to each set of specified comparisons of interest among the treatment groups. For this purpose, let $\boldsymbol{G_C}$ in (7) denote a specified set of $c \leqslant s$ linear contrasts among the $(s+1)$ treatment groups

$$\boldsymbol{G_C} = \begin{bmatrix} \boldsymbol{c}_{11}\boldsymbol{F}_1 + \boldsymbol{c}_{12}\boldsymbol{F}_2 + \cdots + \boldsymbol{c}_{1,s}\boldsymbol{F}_s \\ \cdots \quad \cdots \quad \cdots \quad \cdots \\ \boldsymbol{c}_{c1}\boldsymbol{F}_1 + \boldsymbol{c}_{c2}\boldsymbol{F}_2 + \cdots + \boldsymbol{c}_{c,s}\boldsymbol{F}_s \end{bmatrix}$$

$$= [C \otimes \boldsymbol{I}_{(r+m)}]\boldsymbol{F} \tag{7}$$

with the corresponding matrix of coefficients C having full rank c. In (7), one can note that the $\{c_{li}\}$ are coefficients of the $\bar{\boldsymbol{f}}_i$ and that the $\{-\sum_{i=1}^{s} c_{li}\}$ are coefficients of $\bar{\boldsymbol{f}}_0$.

Non-parametric covariance adjustment is applied to $\boldsymbol{G}_C$ through the linear model

$$\boldsymbol{G}_C \mathrel{\hat{=}} \left\{ \boldsymbol{I}_c \otimes \begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0}_{mr} \end{bmatrix} \right\} \boldsymbol{g}_C = \boldsymbol{X}_C \boldsymbol{g}_C \tag{8}$$

In (8), $\boldsymbol{g}_C$ represents the adjusted counterparts of the contrasts $[C \otimes \boldsymbol{I}_r]\boldsymbol{d}$, where $\boldsymbol{d}=(\boldsymbol{d}_1', \ldots, \boldsymbol{d}_s')'$, among the treatment groups for the means of the response variables. The adjustment is through the equating of the contrasts $[C \otimes \boldsymbol{I}_m]\boldsymbol{u}$, where $\boldsymbol{u}=(\boldsymbol{u}_1', \ldots, \boldsymbol{u}_s')'$, to $\boldsymbol{0}$'s by (8). Determination of $\boldsymbol{g}_C$ is through weighted least squares with weights based on the inverse of the estimated covariance matrix $\boldsymbol{V}_{G,C}$ for $\boldsymbol{G}_C$; that is

$$\boldsymbol{g}_C = [\boldsymbol{X}_C' \boldsymbol{V}_{G,C}^{-1} \boldsymbol{X}_C]^{-1} \boldsymbol{X}_C' \boldsymbol{V}_{G,C}^{-1} \boldsymbol{G}_C \tag{9}$$

where $\boldsymbol{V}_{G,C} = [C \otimes \boldsymbol{I}_{(r+m)}] \boldsymbol{V}_{\boldsymbol{F}} [C' \otimes \boldsymbol{I}_{(r+m)}]$. A consistent estimator for the covariance matrix of $\boldsymbol{g}_C$ is $\boldsymbol{V}_{\boldsymbol{g},C} = [\boldsymbol{X}_C' \boldsymbol{V}_{G,C}^{-1} \boldsymbol{X}_C]^{-1}$. Sufficiently large overall sample size (for example, $n \geqslant 20c(r+m)$) enables $\boldsymbol{g}_C$ to have an approximately multivariate normal distribution for which the covariance matrix is essentially known through $\boldsymbol{V}_{\boldsymbol{g},C}$. Thus, a test statistic for the hypothesis $H_0: \boldsymbol{g}_C \mathrel{\hat{=}} \boldsymbol{0}$ is provided by $Q(\boldsymbol{g}_C) = \boldsymbol{g}_C' \boldsymbol{V}_{\boldsymbol{g},C}^{-1} \boldsymbol{g}_C$. Under H_0, this statistic approximately has the chi-squared distribution with d.f. $= cr$.

3.2. *Estimate of $\boldsymbol{V}_{\boldsymbol{F}}$ for confidence intervals from $\boldsymbol{g}$*

Under the assumption that the patients in the respective treatment groups are conceptually comparable to independent simple random samples, unbiased estimators for the covariance matrices of the $\bar{\boldsymbol{f}}_i$ are the $\boldsymbol{V}_i$ in (10) for $i = 0, 1, \ldots, s$

$$\begin{aligned} \boldsymbol{V}_i &= \sum_{k=1}^{n_i} (\boldsymbol{f}_{ik} - \bar{\boldsymbol{f}}_i)(\boldsymbol{f}_{ik} - \bar{\boldsymbol{f}}_i)' / n_i(n_i - 1) \\ &= \frac{1}{n_i(n_i-1)} \sum_{k=1}^{n_i} \begin{bmatrix} (\boldsymbol{y}_{ik} - \bar{\boldsymbol{y}}_i)(\boldsymbol{y}_{ik} - \bar{\boldsymbol{y}}_i)' & (\boldsymbol{y}_{ik} - \bar{\boldsymbol{y}}_i)(\boldsymbol{x}_{ik} - \bar{\boldsymbol{x}}_i)' \\ (\boldsymbol{x}_{ik} - \bar{\boldsymbol{x}}_i)(\boldsymbol{y}_{ik} - \bar{\boldsymbol{y}}_i)' & (\boldsymbol{x}_{ik} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ik} - \bar{\boldsymbol{x}}_i)' \end{bmatrix} \end{aligned} \tag{10}$$

Since the $\bar{\boldsymbol{f}}_i$ are statistically independent of one another, an unbiased estimator for the covariance matrix of $\bar{\boldsymbol{f}}$ is the block diagonal matrix $\boldsymbol{V} = \operatorname{diag}[\boldsymbol{V}_i]$ with the $\boldsymbol{V}_i$ as diagonal blocks. It follows from $\boldsymbol{F} = \boldsymbol{L}\bar{\boldsymbol{f}}$ in (3) that $\boldsymbol{V}_{\boldsymbol{F}} = \boldsymbol{L}\boldsymbol{V}\boldsymbol{L}'$ is an unbiased estimator for the covariance matrix of $\boldsymbol{F}$.

When the response variables have continuous distributions and the covariance matrices for the $\boldsymbol{f}_{ik}$ are homogeneous for the respective groups, one can replace the $\boldsymbol{V}_i$ by their pooled counterparts

$$\begin{aligned} \bar{\boldsymbol{V}}_i &= \left\{ \sum_{i'=0}^{s} n_{i'}(n_{i'} - 1) \boldsymbol{V}_{i'} \Big/ \sum_{i'=0}^{s} n_i(n_{i'} - 1) \right\} \\ &= \bar{\boldsymbol{V}} / n_i \end{aligned} \tag{11}$$

The estimated covariance matrix $\bar{\boldsymbol{V}}_{\boldsymbol{F}}$ for $\boldsymbol{F}$ then has the structure

$$\bar{\boldsymbol{V}}_{\boldsymbol{F}} = \{[\boldsymbol{L}_1 \boldsymbol{D}_{\phi}^{-1} \boldsymbol{L}_1'] \otimes \bar{\boldsymbol{V}}\}/n \tag{12}$$

where $\boldsymbol{\phi} = (\phi_0, \phi_1, \ldots, \phi_s)' = (n_0, n_1, \ldots, n_s)/n$. Since $\bar{\boldsymbol{V}}$ is based on the overall sample size n, the adjusted estimators $\boldsymbol{g}$ based on it have better compatibility with approximately multivariate normal distributions than their counterparts based on $\boldsymbol{V}_{\boldsymbol{F}}$ for situations where the sample sizes for the separate groups are not clearly large.

3.3. Estimate of $\boldsymbol{V}_{\boldsymbol{F}}$ under the hypothesis of equal treatment effects

The global null hypothesis H_0 of no differences among the effects of the $(s+1)$ treatments for each patient (that is, each patient would have the same vector of responses (and covariables) regardless of the assigned treatment) implies a known structure $\boldsymbol{V}_{\boldsymbol{F},0}$ for the covariance matrix of $\boldsymbol{F}$ [2, 27]. This structure is based on the randomization distribution of $\boldsymbol{F}$ across all possible assignments of the $n = \sum_{i=0}^{s} n_i$ patients to the $(s+1)$ treatments with allocation proportions $\boldsymbol{\phi}$. The known and exact structure for $\boldsymbol{V}_{\boldsymbol{F},0}$ is shown in (13):

$$\boldsymbol{V}_{\boldsymbol{F},0} = \{[\boldsymbol{L}_1 \boldsymbol{D}_{\phi}^{-1} \boldsymbol{L}_1'] \otimes \boldsymbol{V}_0\}/n$$

where

$$\boldsymbol{V}_0 = \frac{1}{n-1} \sum_{i=0}^{s} \sum_{k=1}^{n_i} (\boldsymbol{f}_{ik} - \bar{\boldsymbol{f}})(\boldsymbol{f}_{ik} - \bar{\boldsymbol{f}})' \quad \text{with } \bar{\boldsymbol{f}} = \sum_{i=0}^{s} \phi_i \bar{\boldsymbol{f}}_i \tag{13}$$

The covariance matrix $\boldsymbol{V}_{\boldsymbol{F},0}$ is a matrix of known constants (rather than random variables), and under the hypothesis of equal treatment effects, it is the exact covariance matrix of $\boldsymbol{F}$ relative to its permutation distribution across all possible randomized assignments of patients to the $(s+1)$ treatments. A further advantage of $\boldsymbol{V}_{\boldsymbol{F},0}$ is that it is based on the overall sample size n through $\boldsymbol{V}_0$, and so test statistics such as $Q(\boldsymbol{Kg})$ in (5) with $\boldsymbol{V}_{\boldsymbol{F},0}$ have better compatibility with chi-squared approximations than their counterparts based on $\boldsymbol{V}_{\boldsymbol{F}}$ for situations where the sample sizes for the separate groups are not clearly large. Also, with the use of $\boldsymbol{V}_{\boldsymbol{F},0}$, exact p-values for $Q(\boldsymbol{Kg})$ as a test of H_0 are able to have determination from the corresponding permutation distribution.

4. CLINICAL TRIAL EXAMPLES

4.1. Example 1

The first example for illustrating the methods in this paper has a response surface design to compare combinations of three doses of hydrochlorathiazide (HCTZ) and four doses of an angiotensin converting enzyme inhibitor (ACEI) for blood pressure reduction [9]. There were 12 treatment groups in a 3×4 factorial design with 'placebo' as the lowest level of each factor (HCTZ: H1 = low dose, H2 = high dose; ACEI: A1 = low dose, A2 = medium dose, A3 = high dose), so there were six combinations of HCTZ and ACEI, and about 43 patients per treatment. The responses of interest were changes in supine diastolic and systolic blood pressure and their average change at the last clinic visit relative to baseline, and the covariables

Table I. Results from unadjusted and non-parametric covariance adjusted contrasts among treatments for average of diastolic and systolic blood pressure change from baseline.

Comparison*	Covariance matrix	Scope	Method	Estimate†	Standard error†	*P*-value
(H1A1+H1A2+H2A1+H2A2 $-2\times$H1 $-2\times$H2)	Unequal via $\boldsymbol{V}_i$	Subset	Adjusted	−4.09	1.28	0.001
		All	Adjusted	−4.33	1.24	<0.001
		All	Unadjusted	−4.90	1.42	<0.001
	Equal via $\bar{\boldsymbol{V}}_i$	All	Adjusted	−4.25	1.32	0.001
		All	Unadjusted	−4.90	1.44	<0.001
(H1A1+H1A2+H2A1+H2A2 $-2\times$A1 $-2\times$A2)	Unequal via $\boldsymbol{V}_i$	Subset	Adjusted	−5.14	1.23	<0.001
		All	Adjusted	−5.08	1.22	<0.001
		All	Unadjusted	−5.57	1.31	<0.001
	Equal via $\bar{\boldsymbol{V}}_i$	All	Adjusted	−5.16	1.29	<0.001
		All	Unadjusted	−5.57	1.40	<0.001

*H1 and H2 are low and high doses of HCTZ, and A1, A2, A3 are low, middle and high doses of ACEI.
†The estimate and standard error reported for each comparison is the *average* estimate which involves dividing each of the respective contrasts by 4.

were baseline diastolic and systolic blood pressure and body weight. The design is shown in (14):

	ACEI dose			
HCTZ dose	P	A1	A2	A3
	H1	H1A1	H1A2	H1A3
	H2	H2A1	H2A2	H2A3

(14)

The scope of the analyses includes several components. First, the issue of multiplicity for response variables is addressed by having the average of diastolic and systolic blood pressure change as the primary response for statistical purposes; in this way, significance for it enables separate assessments for diastolic and systolic blood pressures without modification of criteria for significance. Multiplicity for the extent of treatment comparisons is addressed by the evaluation of contrasts for two poolings of four combinations of drug versus their corresponding monotherapies. For example, the average for four treatment groups (H1A1, H1A2, H2A1, H2A2) was compared to both the (H1, H2) average and the (A1, A2) average (with significance for both required to justify combinations); and the average for (H1A2, H1A3, H2A2, H2A3) was compared to the (H1, H2) average and the (A2, A3) average (with significance for both required). If statistical significance applies to both components of one of those contrasts, then the separate combinations within them are compared to the correspondingly pooled monotherapies. The methods for these contrasts among the treatment groups include unadjusted comparisons, adjusted comparisons with respect to the three covariables, and covariance matrix estimation which allows for either unequal covariance matrices across treatments via $\boldsymbol{V}_i$, or homogeneity is assumed across treatments via $\bar{\boldsymbol{V}}_i$ (see Section 3.2 for details).

Table I focuses on the average of systolic and diastolic blood pressure change from baseline as the response of interest. Unadjusted and covariate adjusted estimates are reported for the

differences in average blood pressure change between the combination treatments (averaging of four treatment groups) and their monotherapy counterparts (averaging of two treatment groups). A negative estimate indicates a larger average blood pressure reduction for combination treatments in comparison to the corresponding monotheraphy. The top half of Table I pertains to the comparison with the corresponding HCTZ monotheraphy, while the lower half pertains to the comparison with the ACEI monotherapy.

The first row of Table I provides the difference in average blood pressure change between the combination treatments and the HCTZ monotherapy groups adjusted for the three covariates. Because fewer than 50 patients have random assignment to each of the 12 treatment groups, there may be some concern about potential singularities in the full covariance matrix. A strategy to address this concern is to construct contrasts (as in Section 3.1) for comparisons between averages of specified treatment groups (that is, H1A1, H1A2, H2A1, H2A2 as one group and H1, H2 as the second group) for the response variable and the covariables. Covariate adjustment is then performed on the resulting (4×1) vector of contrasts and (4×4) covariance matrix with such reduced dimensions (see Section 3.1 for details). This strategy is identified as a 'Subset' analysis under the column heading of 'Scope' because covariate adjustment is performed on the reduced data structure for the specified contrasts.

The second row of Table I provides the covariate adjusted estimate of the difference between the change in the blood pressure average for the combination treatments and the average for the HCTZ monotheraphy groups when covariate adjustment is initially performed for all of the 11 treatment groups relative to placebo, and then the resulting adjusted estimates are combined in the contrasts for the groups of interest. In this way, the comparison between the combination treatments and the HCTZ monotheraphy groups has adjustment for equality of the means of the covariables for all 12 groups. The results reported in the second row of Table I are based on this method for estimating the contrast. The third row of Table I provides the corresponding unadjusted comparison for mean change in blood pressure between the combination treatment. The unadjusted estimate is -4.90, indicating that the combination treatments have an average of 4.90 mmHg greater reduction in blood pressure (mean of systolic and diastolic) relative to the average change for HCTZ monotherapy. The standard error for this estimate is 1.42, and so the statistical test for this comparison is statistically significant ($p < 0.001$). The covariate adjusted estimate (which constrains the differences between treatment groups for means of covariates to be null) has a value of -4.33, indicating a slightly smaller effect for the combination treatments relative to the unadjusted estimate, but more importantly the standard error of 1.24 is 13 per cent smaller than its unadjusted counterpart. The resulting p-value is nearly the same with covariate adjustment ($p = 0.001$) as without it. The first three rows of Table I use $\boldsymbol{V}_i$ for estimating the covariance matrix (see Section 3.2), and so has separate estimation of the covariance matrices within each treatment group. This specification allows for potential heterogeneity of the $\boldsymbol{V}_i$.

Rows 4 and 5 of Table I provide the same types of adjusted and unadjusted treatment comparisons as in rows 2 and 3, but with the modification that the covariance matrix used for these results is based on the $\bar{\boldsymbol{V}}_i$ under the assumption that the covariance matrices for the respective treatment groups are homogeneous (see Section 3.2). As with the results previously noted using $\boldsymbol{V}_i$, the covariate adjusted estimate with the $\bar{\boldsymbol{V}}_i$ has a slightly smaller parameter estimate relative to its unadjusted counterpart as a result of adjusting for random imbalance of covariates ($\beta_{\text{unadj}} = -4.90, \beta_{\text{adj}} = -4.25$), and the standard error is smaller by 8 per cent ($\text{SE}_{\text{unadj}} = 1.44, \text{SE}_{\text{adj}} = 1.32$).

Table II. *P*-values from treatment comparisons for response surface clinical trial for reduction of blood pressure relative to baseline.

Comparison*	Criterion	Unadjusted	Adjusted†
(H1A2+H1A3+H2A2+H2A3 $-2\times$H1 $-2\times$H2)	Diastolic	<0.0001	<0.0001
	Systolic	<0.0001	<0.0001
	Average	<0.0001	<0.0001
(H1A2+H1A3+H2A2+H2A3 $-2\times$A2 $-2\times$A3)	Diastolic	<0.0001	<0.0001
	Systolic	0.0001	0.0001
	Average	<0.0001	<0.0001
(H1A1+H1A2+H2A1+H2A2 $-2\times$H1 $-2\times$H2)	Diastolic	0.013	0.012
	Systolic	0.0005	0.0003
	Average	0.0006	0.0005
(H1A1+H1A2+H2A1+H2A2 $-2\times$A1 $-2\times$A2)	Diastolic	<0.0001	0.0001
	Systolic	0.0002	0.0001
	Average	<0.0001	<0.0001
H1A2 − H1	Diastolic	0.003	0.005
	Systolic	0.001	0.002
	Average	0.0006	0.001
H1A2 − A2	Diastolic	0.001	0.004
	Systolic	0.035	0.064
	Average	0.006	0.016

*H1 and H2 are low and high doses of HCTZ, and A1, A2, A3 are low, middle and high doses of ACEI.

†Non-parametric covariance adjusted for baseline systolic blood pressure, baseline diastolic pressure and body weight, using the covariance matrix $\boldsymbol{V}_i$.

The lower half of Table I provides comparisons between the same average of the combination treatments as the upper half of the table and the average of the ACEI monotherapy groups (that is, A1 and A2). For this comparison, covariate adjustment also slightly decreases the treatment parameter estimates, and so this results from constraining the covariate means to be equal for the respective groups. However, the standard error reduction (8 per cent) from covariate adjustment has approximately the same magnitude as the decrease in the treatment parameter estimate (9 per cent). Therefore, in this particular case, the unadjusted and covariate adjusted methods provide very similar *p*-values for the treatment comparison because the reduction in standard error and the treatment parameter estimate with covariance adjustment nearly cancel each other. Table I indicates that the average reduction in blood pressure relative to baseline for the four specified combination treatments is significantly greater than either of the respective monotherapies (ACEI or HCTZ), and non-parametric covariance adjustment clarifies these results by adjusting for the effect of random covariate imbalance between the treatment groups and by reducing the variance of the treatment comparison estimate.

In Table II, three different responses of interest are evaluated: change in systolic blood pressure; change in diastolic blood pressure; and the average of the two using the estimated $\boldsymbol{V}_i$ covariance matrix. In this table, only *p*-values are reported. In the first section of Table II, the average for the H1A2, H1A3, H2A2 and H2A3 combinations is compared to its two monotherapy counterparts (that is, the H1, H2 average and the A2, A3 average, respectively).

Here, the average for the combination treatments is for the two highest doses of the ACE inhibitor with the two doses of HCTZ. For systolic, diastolic and average blood pressure change, covariate adjustment provides similar p-values relative to its unadjusted counterparts. Although the components are not shown in Table II, covariate adjustment reduces the treatment parameter estimate as a result of constraining differences in covariate means to equal zero, and adjustment also reduces the standard error of this estimate. The net results are similar p-values for both the unadjusted and covariate adjusted tests. In the second portion of Table II, the average for the combination treatments is for the two lower doses of the ACE inhibitor with the two doses of HCTZ. This grouping is separately compared to the two respective averages for the corresponding monotherapies.

By averaging treatment groups in specified contrasts, we address issues regarding multiple comparisons among groups. Because the contrasts for averages were significant (for example, two-sided $p<0.025$) in the top two portions of Table II, there is reasonable justification for comparisons of each of the separate combination treatments to their corresponding monotherapies. Results for these types of comparisons are illustrated in the bottom third of Table II for comparisons between H1A2 (low HCTZ and medium ACEI) to H1 and A2. The combination treatment H1A2 has significantly reduced diastolic, systolic and average blood pressure relative to the low dose of HCTZ alone. For the comparison of the combination H1A2 with its A2 monotherapy, covariate adjustment actually increases the p-value for the test of treatment comparison ($p_{\text{unadj}}=0.035$, $p_{\text{adj}}=0.064$). If we had modelled last clinic visit systolic blood pressure as the response of interest (instead of the change from baseline), the unadjusted p-value for the comparison of the two treatment groups would be $p_{\text{unadj}}=0.118$. The covariate adjusted $p_{\text{adj}}=0.064$ is the same result one would get whether modelling the *change* in SBP or the *last visit* SBP as the outcome. It is possible that the unadjusted p-value for blood pressure change ($p_{\text{unadj}}=0.035$) overstates the strength of this treatment comparison because of random imbalances in covariate means. By forcing differences in covariate means to equal zero, the covariate adjusted method probably provides a more realistic result.

4.2. Example 2

The second example is a study to compare three doses of test drug to placebo for lengthening time to disease progression or death for a fatal disorder. A total of 959 patients were randomized to this trial providing approximately 240 subjects per treatment group [6, 8]. The primary comparison for this trial is the middle dose versus placebo. A supportive overall comparison is the pooling of the three doses versus placebo. Both the Wilcoxon and logrank tests are applied in parallel. However, the Wilcoxon test would be better as a primary test in this case because of its ability to detect test drug's reduction of early deaths or progression [32]. The differences between treatment groups are more evident at intermediate points of follow-up rather than at the end (that is, 12 months instead of 18 months). This can be seen in Table III where the event rates for test treatment relative to placebo appear to be more beneficial at the earlier time point. If there are reasons for expecting treatment benefit to increase until 12 months and then become somewhat less after that time, then it would be better to prespecify this intermediate region as the primary interval of interest for analyses. In the analysis plan for this study, there are 24 covariates with *a priori* identification. Since most of these covariables are expected to be strongly correlated with response, it is better for the covariate adjusted analyses to have the primary role rather than the unadjusted analyses.

Table III. Randomized clinical trial to compare three doses of test drug and placebo for time to disease progression.

Treatment	Sample size	Event by 12 months	Event by 18 months
Placebo	242	37.2%	49.6%
Low dose	237	29.5%	44.7%
Middle dose	236	26.3%	43.2%
High dose	244	27.5%	42.2%
Total	959	30.1%	44.9%

Losses to follow-up managed as survivors of event in this display and as censored in others. Only six patients were lost to follow-up prior to 12 months.

The number of covariates is reduced to a subset of 14 in order to have better support for approximate normality from central limit theory through the consideration of a more manageable number of degrees of freedom. Stepwise multiple linear regression for Wilcoxon rank scores, with the effects for treatments not included in the model, is used to identify this subset of 14 covariables.

Analyses for several response criteria are illustrated for this clinical trial. They include percentages with the event by 12 months, means of Wilcoxon rank scores and logrank scores, and averages of cumulative survival rates. These criteria are assessed with unadjusted comparisons among treatments, and adjusted comparisons among treatments with respect to the subset of 14 covariables. Non-parametric methods which specify no differences among groups for means of covariables (on the basis of randomization in the study design) are emphasized for the adjusted comparisons. For supportive and complementary purposes, a logistic regression model for event by 12 months or not and a proportional hazards model for time to event over 18 months are applied.

Table IV provides the results of treatment comparisons with respect to progression or death at 12 months as a dichotomous response (Yes/No). The six individuals with censored times prior to 12 months are managed as progression-free survivors in this example. The top section of Table IV provides the comparison of the average of high and medium doses of test treatment groups to placebo. Having this comparison first can be a good strategy for addressing multiple treatment comparisons since significance for it at the two-sided 0.05 level allows for assessment for the separate comparisons of high dose and medium dose to placebo at the two-sided 0.05 level. On average, patients in the high and medium dose groups have 10.3 per cent fewer events than the placebo group at 12 months since a negative parameter estimate indicates a treatment benefit. After adjustment for the 14 covariates, treatment benefit increases slightly to 10.9 per cent fewer events for test treatment and the standard error is reduced by about 20 per cent from 3.72 per cent to 3.02 per cent, leading to a stronger result for the comparison ($p_{\text{unadj}} = 0.006$, $p_{\text{adj}} < 0.001$). The next three parts of Table IV provide the unadjusted and covariate adjusted results for each of the pairwise comparisons of test treatment and placebo, and the fifth part provides the results of the average of low, medium and high dose groups versus placebo. The primary comparison, medium – placebo, indicates a significant benefit for test treatment which becomes stronger with covariate adjustment ($p_{\text{unadj}} = 0.010$, $p_{\text{adj}} = 0.001$).

It is possible to specify other contrasts for linear hypotheses of the treatment parameters in order to address multiple comparisons. In the second to last part of Table IV, a test of

Table IV. Results from unadjusted and non-parametric covariance adjusted comparisons among treatments for occurrence or not of event by 12 months.

Comparison*	Method†	Estimate	Standard error	*P*-value
H + M − 2P	Adjusted	−10.94%	3.02%	<0.001
	Unadjusted	−10.33%	3.72%	0.006
H − P	Adjusted	−10.54%	3.44%	0.002
	Unadjusted	−9.73%	4.23%	0.021
M − P	Adjusted	−11.35%	3.52%	0.001
	Unadjusted	−10.92%	4.23%	0.010
L − P	Adjusted	−7.70%	3.54%	0.033
	Unadjusted	−7.65%	4.30%	0.075
H + M + L − 3P	Adjusted	−9.76%	2.82%	<0.001
	Unadjusted	−9.43%	3.53%	0.008
3H + M − L − 3P	Adjusted	Not applicable		0.001
	Unadjusted	Not applicable		0.015
(H, M, L, P)	Adjusted	Not applicable		0.004
	Unadjusted	Not applicable		0.050

*H = high dose, M = middle dose, L = low dose, P= placebo. Non-parametric method used $\boldsymbol{V}_F$ in estimation.
$Q = 30.04$ (d.f. = 42, $p = 0.916$) for random imbalances.

†For the adjusted method: non-parametric covariance adjusted for 14 covariates identified by stepwise regression from a prespecified set of 24 with $\boldsymbol{V}_i$ as covariance matrix.

the hypothesis for 'no linear trend' is presented. Covariance adjustment in this case provides a stronger result for greater efficacy with increasing dose ($p_{\text{unadj}} = 0.015$, $p_{\text{adj}} = 0.001$). The last section of Table IV pertains to the global null hypothesis of equality of the effects of all four treatments (that is, the respective treatment parameters for the differences between each of the doses and placebo are jointly null). The corresponding Wald chi-square statistic has 3 degrees of freedom, and results from both the unadjusted and covariate adjusted tests indicate that at least one of the treatment comparisons is non-null ($p_{\text{unadj}} = 0.050$, $p_{\text{adj}} = 0.004$). See (5) in Section 3 for specifications concerning these types of linear hypotheses.

As described in Section 3, the non-parametric covariate adjustment method involves constraining the differences between treatment groups for covariate means to equal zero, and these constraints follow from the randomization process. It is also possible to evaluate the extent to which a given randomization of patients is atypical by evaluating goodness-of-fit (GOF) by the statistic in (6) of Section 3. The footnote at the bottom of Table IV provides the GOF statistic for evaluating the extent of random imbalance among treatment groups. The Wald chi-square statistic = 30.04 and has 42 degrees of freedom (14 covariates × 3 treatment comparisons with placebo). The resulting p-value ($p = 0.916$) indicates that the random imbalance among the four treatment groups for distributions of covariates is compatible with expectations from randomization, and so constraining the differences among covariate means for treatment groups to equal zero leads to adjusted estimates which are similar to their unadjusted counterparts (that is, they do not differ by more than 25 per cent of the corresponding standard errors).

Because estimated survival rates over time are not typically smooth functions, it is often advantageous to obtain an average estimate over a defined period of time. For the results in Table V, life table estimates [33] are formed at 16, 17 and 18 months, and then the average of

Table V. Results from unadjusted and non-parametric covariance adjusted comparisons among treatments for average differences in survival rates at 16 months, 17 months and 18 months.

Comparison*	Method†	Estimate	Standard error	*P*-value
H + M − 2P	Adjusted	8.17%	3.09%	0.008
	Unadjusted	7.25%	3.92%	0.064
H − P	Adjusted	6.95%	3.61%	0.054
	Unadjusted	5.87%	4.54%	0.196
M − P	Adjusted	9.38%	3.61%	0.009
	Unadjusted	8.62%	4.50%	0.055
L − P	Adjusted	4.49%	3.71%	0.226
	Unadjusted	4.99%	4.54%	0.273
H + M + L − 3P	Adjusted	6.93%	2.90%	0.017
	Unadjusted	6.49%	3.69%	0.079
3H + M − L − 3P	Adjusted		Not Applicable	0.025
	Unadjusted		Not Applicable	0.139
(H, M, L, P)	Adjusted		Not Applicable	0.059
	Unadjusted		Not Applicable	0.281

*H = high dose, M = middle dose, L = low dose, P = Placebo. Non-parametric method used $\boldsymbol{V_F}$ in estimation. $Q = 6.08$ (with d.f. = 6, $p = 0.414$) for homogeneity across intervals in the covariate adjusted case.

†For the adjusted method: non-parametric covariance adjusted for 14 covariates identified by stepwise regression from a prespecified set of 24 with $\boldsymbol{V_{R,i}}$ as covariance matrix (see Appendix A2).

the estimates for these three time points ('smoothed' estimate) is used. The response criteria of interest are the differences in these average survival rates between treatment groups in a sense analogous to the area between the Kaplan–Meier curves during 16–18 months [6]. Because survival rates involve a product of ratios of means (that is, the numerators indicate the proportions of patients without the event of interest, and the denominators indicate the proportions at risk for the event during successive intervals), we need to take their structure as functions of means into account when applying non-parametric methods for covariance adjustment (see Appendix A2 and Tangen and Koch [6]).

For the two primary treatment comparisons in Table V, we see that covariate adjustment provides substantial improvement in power. The unadjusted medium – placebo comparison indicates that 8.6 per cent fewer patients have progressed by the 16–18 month interval relative to placebo, and this progression-free difference approaches significance ($p = 0.055$). By constraining the differences between the two treatment groups for means of the 14 covariates to be zero, we obtain a covariate adjusted estimate which is slightly larger (9.4 per cent progression-free difference), and more importantly we see a reduction in the standard error of 20 per cent, leading to a stronger result for the comparison ($p = 0.009$). A similar pattern is also seen for the pooled comparison of active treatment (L, M, H) relative to placebo ($p_{\text{unadj}} = 0.079$, $p_{\text{adj}} = 0.017$). Tests of hypotheses pertaining to linear trend and global treatment differences also benefit from covariance adjustment through stronger results.

Because we are averaging three interval-specific parameter estimates (16, 17 and 18 months) in Table V, it is reasonable to evaluate whether there is homogeneity of treatment differences among these three estimates. The footnote at the bottom of Table V provides a Wald chi-square = 6.08 with 6 degrees of freedom ((number of intervals of time − 1 = 2) × 3 treatment comparisons) and $p = 0.414$, supporting the averaging ('smoothing') of the treatment comparisons across the three specified intervals.

Table VI. *P*-values from treatment comparisons for clinical trial with evaluation of survival.

Comparison*	Statistic	Unadjusted	Adjusted†
(H + M − 2P)	Logrank	0.050	0.014
	Wilcoxon	0.032	0.006
(H − P)	Logrank	0.092	0.025
	Wilcoxon	0.077	0.039
(M − P)	Logrank	0.087	0.043
	Wilcoxon	0.053	0.020
(L − P)	Logrank	0.251	0.251
	Wilcoxon	0.209	0.190
(H + M + L − 3P)	Logrank	0.063	0.027
	Wilcoxon	0.043	0.013
(3H + M − L − 3P)	Logrank	0.076	0.016
	Wilcoxon	0.059	0.009
(H, M, L, P)	Logrank	0.278	0.101
	Wilcoxon	0.203	0.055

*H = high dose, M = middle dose, L = low dose, P = placebo.
†Non-parametric covariance adjusted for 14 covariates identified by stepwise regression from a prespecified set of 24 with $\boldsymbol{V}_{\boldsymbol{F},0}$ as covariance matrix. Non-parametric method had 20 per cent smaller standard error.

Table VI provides unadjusted and covariate adjusted *p*-values from logrank and Wilcoxon tests [34, 35]. Time-to-event scores are assigned to all patients in the study, and mean scores are constructed for each treatment group. The *p*-values provided in this table are the result of testing whether the mean logrank (or Wilcoxon) scores for all treatment groups are equal. Because we are in the hypothesis testing setting, the covariance matrix $\boldsymbol{V}_{\boldsymbol{F},0}$ is used instead of $\boldsymbol{V}_{\boldsymbol{F}}$ (see Section 3.3 and Koch *et al.* [2] for more details) to test the null hypothesis of equal treatment effects. As stated earlier in the paper for this clinical trial, the Wilcoxon test tends to provide a stronger result for each comparison of treatment effects. In agreement with Tables IV and V, covariance adjustment provides stronger results for tests concerning treatment effects in Table VI, and although not shown, the standard errors in the covariate adjusted tests are about 20 per cent smaller than their unadjusted counterparts.

The results from regression model based methods are shown in Table VII for the logistic regression model with disease progression at 12 months (Yes versus No) as the response of interest and the proportional hazards (PH) regression model with time to disease progression over the 18 month interval as the response of interest [36]. The unadjusted model includes three indicators for the three doses of test treatment relative to placebo, and the covariate adjusted model includes the 14 covariates in addition to the three indicators for treatment effects. The unadjusted parameter estimate for each model is a population average log-odds ratio or log-hazard ratio, respectively. The covariate adjusted parameter, on the other hand, has a patient-specific nature by applying to individuals who share the same covariate values that have conditioning in the model. For the primary comparison of medium versus placebo, the negative parameter estimates indicate a test treatment benefit in both the logistic regression and PH models. However, conditioning on the 14 covariates leads to substantially larger treatment parameter estimates (35 per cent increase in logistic regression model and 66 per cent increase for the PH model, with both exceeding the corresponding errors), and the standard errors actually increase slightly [37, 38]. However, the increase in the treatment parameter estimate is relatively bigger than the increase in the standard error, and so the net results

Table VII. Results from logistic regression model for event by 12 months and proportional hazards regression model for time to event over 18 months.

Comparison*	Model	Method	Estimate	Standard error	*P*-value
H − P	Logistic regression	Adjusted	−0.719	0.242	0.003
		Unadjusted	−0.447	0.196	0.022
	Proportional hazards	Adjusted	−0.446	0.138	0.001
		Unadjusted	−0.226	0.134	0.093
M − P	Logistic regression	Adjusted	−0.687	0.242	0.005
		Unadjusted	−0.508	0.199	0.011
	Proportional hazards	Adjusted	−0.382	0.137	0.005
		Unadjusted	−0.230	0.135	0.088
L − P	Logistic regression	Adjusted	−0.450	0.240	0.061
		Unadjusted	−0.346	0.195	0.076
	Proportional hazards	Adjusted	−0.226	0.135	0.095
		Unadjusted	−0.153	0.133	0.252

*H = high dose, M = middle dose, L = low dose, P = placebo.
Adjusted analysis accounted for 14 covariables identified by stepwise regression (from a prespecified set of 24) for Wilcoxon rank scores.
For (H + M − 2P), unadjusted $p = 0.005$ and adjusted $p = 0.001$ in logistic regression model, and unadjusted $p = 0.048$ and adjusted $p < 0.001$ in proportional hazards model.

are more significant tests concerning treatment effects (logistic: $p_{\text{unadj}} = 0.011$, $p_{\text{adj}} = 0.005$; PH: $p_{\text{unadj}} = 0.088$, $p_{\text{adj}} = 0.005$ for medium versus placebo). For the most part, the regression model based p-values in Table VII are in agreement with their non-parametric counterparts in Tables IV–VI, and this similarity reinforces the greater usefulness of both methods than the unadjusted analysis without covariate adjustment.

5. DISCUSSION

In confirmatory clinical trials that are designed to compare multiple doses of test treatment to a control, there are issues regarding compound hypotheses and multiple comparisons that need to be considered, and so controlling the experimentwise significance level is a concern. In most cases, the analysis plan needs to prespecify clearly the proposed order of tests to be conducted (and/or how the overall significance level will be allocated among multiple tests), which statistical method will be used, and whether adjustment for covariates will be performed.

There are several benefits from specifying non-parametric ANCOVA for performing the primary analyses. No or only minimal assumptions are needed beyond randomization in the study design, whereas regression model based methods have assumptions about model fit for which departures may require modifications that are incompatible with a fully prespecified analysis plan [39]. However, if the intent of the clinical trial examples in this paper were dose-finding rather than confirmatory, parametric regression modelling might be more useful since *post hoc* refinement of the models would not be a major issue. Non-parametric methods provide traditionally expected results of analysis of covariance; namely, a typically small adjustment to the estimate of a treatment comparison (so as to account for random imbalance of covariates between treatment groups) and variance reduction for this estimate when covariates

are strongly correlated with the response of interest. By using non-parametric methods, one can obtain an increase in power for treatment comparisons without concern for modelling assumptions with uncertain applicability. In a reinforcing way, statistical regression models can provide useful information for the relationship of response variables to covariables and treatment for a clinical trial, and in this complementary way, their supportive results can be helpful for purposes of generalizability [2, 6–8].

APPENDIX: ADDITIONAL ASPECTS OF NON-PARAMETRIC ANCOVA

A1. Extensions with adjustment for stratification

Let $h = 1, 2, \ldots, q$ index a set of strata within which patients are randomized to the $(s+1)$ treatment groups. Let $\boldsymbol{F}_h$, $\boldsymbol{V}_{\boldsymbol{F}_h}$ and n_{h+} denote the counterparts for the hth stratum to $\boldsymbol{F}, \boldsymbol{V}_{\boldsymbol{F}}$ and n. For situations with large sample sizes within each stratum, one first applies covariance adjustment within each stratum to obtain $\boldsymbol{g}_h = [\boldsymbol{X}' \boldsymbol{V}_{\boldsymbol{F}_h}^{-1} \boldsymbol{X}]^{-1} \boldsymbol{X}' \boldsymbol{V}_{\boldsymbol{F}_h}^{-1} \boldsymbol{F}_h$ and $\boldsymbol{V}_{\boldsymbol{g}_h} = [\boldsymbol{X}' \boldsymbol{V}_{\boldsymbol{F}_h}^{-1} \boldsymbol{X}]^{-1}$ where $\boldsymbol{X}$ has the same form as in (3). Adjustment for stratification can then be applied through the determination of the weighted mean $\bar{\boldsymbol{g}} = \sum_{h=1}^{q} w_h \boldsymbol{g}_h$ and its estimated covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{g}}} = \sum_{h=1}^{q} w_h^2 \boldsymbol{V}_{\boldsymbol{g}_h}$ where $0 \leqslant w_h \leqslant 1$ and $\sum_{h=1}^{q} w_h = 1$; usually $w_h = (n_{h+} / \sum_{h'=1}^{q} n_{h'+})$, although other weights are possible. One can then use $\bar{\boldsymbol{g}}$ and $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ for hypothesis tests and confidence intervals in ways like those previously described for $\boldsymbol{g}$ and $\boldsymbol{V}_{\boldsymbol{g}}$.

When sample sizes are not large enough to support covariance adjustment within each stratum (because of near or actual singularities in the $\boldsymbol{V}_{\boldsymbol{F}_h}$), adjustment for stratification is applied first through the determination of $\boldsymbol{F}_w = \sum_{h=1}^{q} w_h \boldsymbol{F}_h$ and $\boldsymbol{V}_{\boldsymbol{F}_w} = \sum_{h=1}^{q} w_h^2 \boldsymbol{V}_{\boldsymbol{F}_h}$. Non-parametric covariance adjustment is then applied to $\boldsymbol{F}_w$ through the model (3) for $\boldsymbol{F}_w$. The results from weighted least squares are $\boldsymbol{g}_w = [\boldsymbol{X}' \boldsymbol{V}_{\boldsymbol{F}_w}^{-1} \boldsymbol{X}]^{-1} \boldsymbol{X}' \boldsymbol{V}_{\boldsymbol{F}_w}^{-1} \boldsymbol{F}_w$ and its estimated covariance matrix $\boldsymbol{V}_{\boldsymbol{g}_w} = [\boldsymbol{X}' \boldsymbol{V}_{\boldsymbol{F}_w}^{-1} \boldsymbol{X}]^{-1}$. These quantities can then be used for hypothesis tests and confidence intervals in ways like those previously described for $\boldsymbol{g}$ and $\boldsymbol{V}_{\boldsymbol{g}}$.

A2. Extensions to account for compound linear and multiplicative functions of response means

In some applications, compound linear and multiplicative functions of the $\bar{\boldsymbol{y}}_i$ are of interest for covariance adjustment. The scope of such functions includes odds ratios for dichotomous and ordinal outcomes [4] as well as incidence density ratios and cumulative survival rates for times to events [6, 8]. Analysis of these functions (or their logarithms) involves their determination as $\boldsymbol{a}(\bar{\boldsymbol{y}}_i)$ in the transformations $\boldsymbol{R}_i = \boldsymbol{R}(\bar{\boldsymbol{f}}_i) = (\boldsymbol{a}(\bar{\boldsymbol{y}}_i)', \bar{\boldsymbol{x}}_i')'$ of the $\bar{\boldsymbol{f}}_i$ for the respective groups. Also, with a as the dimension of $\boldsymbol{a}(\bar{\boldsymbol{y}}_i)$, the dimension of $\boldsymbol{R}_i$ is $(a+m)$. Through linear Taylor series methods [reference 27, Chapter 12], a consistent estimate for the covariance matrix of $\boldsymbol{R}_i$ is $\boldsymbol{V}_{\boldsymbol{R},i}$ in (A1):

$$\boldsymbol{V}_{\boldsymbol{R},i} = \begin{bmatrix} \boldsymbol{A}_i(\bar{\boldsymbol{y}}_i) & \boldsymbol{0}_{am} \\ \boldsymbol{0}_{ma} & \boldsymbol{I} \end{bmatrix} \boldsymbol{V}_i \begin{bmatrix} \boldsymbol{A}_i(\bar{\boldsymbol{y}}_i) & \boldsymbol{0}_{am} \\ \boldsymbol{0}_{ma} & \boldsymbol{I} \end{bmatrix}' \tag{A1}$$

where $\boldsymbol{A}_i(\bar{\boldsymbol{y}}_i) = [\partial \boldsymbol{a}(z) / \partial z | z = \bar{\boldsymbol{y}}_i]$. The $\boldsymbol{R}_i$ and $\boldsymbol{V}_{\boldsymbol{R},i}$ are then managed in the same ways as previously described for the $\bar{\boldsymbol{f}}_i$ and $\boldsymbol{V}_i$. In particular, the concatenated vector $\boldsymbol{R} = (\boldsymbol{R}_0', \ldots, \boldsymbol{R}_s')'$ and its estimated covariance matrix $\boldsymbol{V}_{\boldsymbol{R}} = \text{diag}[\boldsymbol{V}_{\boldsymbol{R},i}]$ are formed. Then non-parametric covariance

adjustment is applied to $\boldsymbol{F_R} = \boldsymbol{LR}$ relative to $\boldsymbol{V_{F_R}} = \boldsymbol{LV_RL'}$ through methods like equations (3)–(6) in Section 3.

A3. Simplified forms for $\boldsymbol{g}$ and $\boldsymbol{V_g}$ is situations where $\boldsymbol{V_F}$ is based on $\bar{\boldsymbol{V}}$ or $\boldsymbol{V}_0$

As indicated in equations (12) and (13) of Sections 3.2 and 3.3, $\boldsymbol{V_F}$ has the simplified form

$$\boldsymbol{V_F} = \{[\boldsymbol{L}_1\boldsymbol{D}_\phi^{-1}\boldsymbol{L}_1'] \otimes \boldsymbol{V}\}/n \tag{A2}$$

where $\boldsymbol{V} = (\bar{\boldsymbol{V}}$ or $\boldsymbol{V}_0)$ in situations where $\bar{\boldsymbol{V}}$ or $\boldsymbol{V}_0$ are applicable. This simplified structure implies that

$$\begin{aligned}\boldsymbol{V_g} &= (\boldsymbol{X'V_F^{-1}X})^{-1} = [\boldsymbol{L}_1\boldsymbol{D}_\phi^{-1}\boldsymbol{L}_1'/n] \otimes \left\{ \begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0}_{mr} \end{bmatrix}' \boldsymbol{V}^{-1} \begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0}_{mr} \end{bmatrix} \right\}^{-1} \\ &= \{\boldsymbol{L}_1\boldsymbol{D}_\phi^{-1}\boldsymbol{L}_1' \otimes [\boldsymbol{V}_{yy} - \boldsymbol{V}_{xy}'\boldsymbol{V}_{xx}^{-1}\boldsymbol{V}_{xy}]\}/n \end{aligned} \tag{A3}$$

where $\boldsymbol{V}_{yy}$, $\boldsymbol{V}_{xy}$, and $\boldsymbol{V}_{xx}$ are the blocks in the partition of $\boldsymbol{V}$ corresponding to the covariance matrix of the $\boldsymbol{y}_{ik}$, the covariances of the $\boldsymbol{x}_{ik}$ with the $\boldsymbol{y}_{ik}$, and the covariance matrix of the $\boldsymbol{x}_{ik}$. From the structure of $\boldsymbol{V_g}$, one can see that the relationship of $(\boldsymbol{V}_{yy} - \boldsymbol{V}_{xy}'\boldsymbol{V}_{xx}^{-1}\boldsymbol{V}_{xy})$ to $\boldsymbol{V}_{yy}$ expresses the extent to which smaller variances apply to assessments based on $\boldsymbol{g}$ than to those based on their unadjusted counterparts $\boldsymbol{d}$. One can further note that

$$\begin{aligned}\boldsymbol{g} &= \boldsymbol{V_gX'V_F^{-1}F} = \{\boldsymbol{I}_s \otimes [\boldsymbol{I}_r, -\boldsymbol{V}_{xy}'\boldsymbol{V}_{xx}^{-1}]\}\boldsymbol{F} \\ &= (\boldsymbol{g}_1', \ldots, \boldsymbol{g}_s')'\end{aligned}$$

with $\boldsymbol{g}_i = (\boldsymbol{d}_i - \boldsymbol{V}_{xy}'\boldsymbol{V}_{xx}^{-1}\boldsymbol{u}_i)$ for $i = 1, 2, \ldots, s$. Thus, the $\boldsymbol{g}_i$ are differences between the ith treatment group and the control (or 0th) group for means of residuals from the ordinary least squares fits of separate multiple linear regression models which have the $\boldsymbol{x}_{ik}$ as explanatory variables for each of the response variables y_{ikj} where $j = 1, 2, \ldots, r$.

In the rarely observed special case when $\boldsymbol{u} = \boldsymbol{0}$, $\boldsymbol{g} = \boldsymbol{d}$, but there is still variance reduction through $\boldsymbol{V_g}$ being the estimated covariance matrix of $\boldsymbol{g}$; also, since $\hat{\boldsymbol{F}} = \boldsymbol{Xg} = \boldsymbol{F}$ in this case, it follows from (6) that $Q = 0$ with $p = 1.00$. The properties noted here for this special case similarly apply when $\boldsymbol{V_F}$ is based on the $\boldsymbol{V}_i$ in (10).

ACKNOWLEDGEMENTS

The authors of this paper thank Stephen Senn and Byron Jones for their critical comments and suggestion on a previous revision of this paper.

REFERENCES

1. Koch GG, Amara IA, Davis GW, Gillings DB. A review of some statistical methods for covariance analysis of categorical data. *Biometrics* 1982; **38**:563–595.
2. Koch GG, Tangen CM, Jung JW, Amara IA. Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine* 1998; **17**:1863–1892.

3. Jung JW, Koch GG. A linear model method for rank measures of association from longitudinal studies with fixed conditions (visits) for data collection and more than two groups. *Journal of Biopharmaceutical Statistics* 1998; **8**(2):299–316.
4. Tangen CM, Koch GG. Complementary nonparametric analysis of covariance for logistic regression in a randomized clinical trial setting. *Journal of Biopharmaceutical Statistics* 1999; **9**(1):45–66.
5. Jung JW, Koch GG. Multivariate non-parametric methods for Mann–Whitney statistics to analyse cross-over studies with two treatment sequences. *Statistics in Medicine* 1999; **18**(8):989–1017.
6. Tangen CM, Koch GG. Nonparametric analysis of covariance for hypothesis testing with logrank and Wilcoxon scores and survival-rate estimation in a randomized clinical trial. *Journal of Biopharmaceutical Statistics* 1999; **9**(2):307–338.
7. Koch GG, Tangen CM. Nonparametric analysis of covariance and its role in non-inferiority clinical trials. *Drug Information Journal* 1999; **33**(4):1145–1159.
8. Tangen CM, Koch GG. Non-parametric covariance methods for incidence density analyses of time-to-event data from a randomized clinical trial and their complementary roles to proportional hazards regression. *Statistics in Medicine* 2000; **19**:1039–1058.
9. Phillips JA, Cairns V, Koch GG. The analysis of a multiple-dose, combination-drug clinical trial using response surface methodology. *Journal of Biopharmaceutical Statistics* 1992; **2**(1):49–67.
10. Koch GG, Davis SM, Anderson RL. Methodological advances and plans for improving regulatory success for confirmatory studies. *Statistics in Medicine* 1998; **17**:1675–1690.
11. Ruberg SG. Contrasts for identifying the minimum effective dose. *Journal of the American Statistical Association* 1989; **84**:816–822.
12. Ruberg SJ. Dose response studies. I. Some design considerations. *Journal of Biopharmaceutical Statistics* 1995; **5**(1):1–14.
13. Ruberg SJ. Dose response studies. II. Analysis and interpretation. *Journal of Biopharmaceutical Statistics* 1995; **5**(1):15–42.
14. Senn SJ. *Statistical Issues in Drug Development*. Wiley: Chichester, 1997.
15. Bauer P. Multiple testing in clinical trials. *Statistics in Medicine* 1991; **10**:871–890.
16. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; **75**:800–802.
17. Hochberg Y, Tamhane AC. *Multiple Comparison Procedures.* Wiley: New York, 1987.
18. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 1988; **75**:383–386.
19. Hsu JC. *Multiple Comparisons*. Chapman and Hall: London, 1996.
20. Koch GG, Gansky SA. Statistical considerations for multiplicity in confirmatory protocols. *Drug Information Journal* 1996; **30**:523–534.
21. Rom DM, Costello RJ, Connell LT. On closed test procedures for dose-response analysis. *Statistics in Medicine* 1994; **13**:1583–1596.
22. Sarkar S, Chang CK. The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association* 1997; **92**:1601–1608.
23. Tamhane AC, Hochberg Y, Dunnett CW. Multiple test procedures for dose finding. *Biometrics* 1996; **52**: 21–37.
24. Westfall PH, Tobias RD, Rom D, Wolfinger RD, Hochberg Y. *Multiple Comparisons and Multiple Tests using the SAS System*. SAS Institute Inc: Cary, NC, 1999.
25. Westfall PH, Young SS. *Resampling-based Multiple Testing*: *Examples and Methods for p-value Adjustment*. Wiley: New York, 1993.
26. Altman DG, Dore CJ. Letters to the editor: Baseline comparisons in randomized clinical trials. *Statistics in Medicine* 1991; **10**:797–802.
27. Stokes ME, Davis CS, Koch GG. *Categorical Data Analysis using the SAS System*. SAS Institue Inc: Cary, NC, 1995.
28. Chastang C, Byar D, Piantadosi S. A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models. *Statistics in Medicine* 1988; **7**: 1243–1255.
29. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 1984; **71**(3):431–444.
30. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials* 1998; **19**:249–256.
31. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 1989; **8**:467–475.
32. Lee ET. *Statistical Methods for Survival Data Analysis*. Wiley: New York, 1992.
33. Koch GG, Johnson WD, Tolley HD. A linear models approach to the analysis of survival and extent of disease in multidimensional contingency tables. *Journal of the American Statistical Association* 1972; **67**: 783–796.
34. Peto R, Peto J. Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statistical Society*, *Series B* 1972; **34**:205–207.

35. Prentice RL. Linear rank tests with right censored data. *Biometrika* 1978; **65**:167–179.
36. Collett D. *Modeling Survival Data in Medical Research.* Chapman and Hall: London, 1994.
37. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review* 1991; **58**(2):227–240.
38. Ford I, Norrie J, Atimadis S. Model inconsistency, illustrated by the Cox proportional hazards model. *Statistics in Medicine* 1995; **14**:735–746.
39. Lewis JA, Jones DR, Röhmel J. Biostatistical methodology in clinical trials—a European Guideline. *Statistics in Medicine* 1995; **14**:1655–1682.

