Group sequential crossover trial designs with strong control of the familywise error rate

Michael J. Grayling, James M. S. Wason & Adrian P. Mander

Published online: 02 Oct 2018.

Submit your article to this journal

Article views: 32

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Group sequential crossover trial designs with strong control of the familywise error rate

Michael J. Grayling, James M. S. Wason, and Adrian P. Mander

Hub for Trials Methodology Research, MRC Biostatistics Unit, Cambridge, UK

ABSTRACT
Crossover designs are an extremely useful tool to investigators, and group sequential methods have proven highly proficient at improving the efficiency of parallel group trials. Yet, group sequential methods and crossover designs have rarely been paired together. One possible explanation for this could be the absence of a formal proof of how to strongly control the familywise error rate in the case when multiple comparisons will be made. Here, we provide this proof, valid for any number of initial experimental treatments and any number of stages, when results are analyzed using a linear mixed model. We then establish formulae for the expected sample size and expected number of observations of such a trial, given any choice of stopping boundaries. Finally, utilizing the four-treatment, four-period TOMADO trial as an example, we demonstrate that group sequential methods in this setting could have reduced the trials expected number of observations under the global null hypothesis by over 33%.

1. Introduction

The efficiency of crossover trials often makes them the best design for a clinical trial. Administering multiple treatments to patients reduces the standard error of the estimated treatment effects compared to a parallel trial design with an equal number of patients. Therefore, though restrictions to their use exist, such as a requirement for patients to begin each new treatment period in a state comparable to those completed, crossover trials are the design of choice in many settings (Jones and Kenward, 2014; Senn, 2002), resulting in them accounting for 22% of all published trials in December 2000, for example (Mills et al., 2009).

In a parallel design setting, group sequential methods are frequently utilized to improve a clinical trial's efficiency (Jennison and Turnbull, 2000). These designs incorporate interim analyses that allow for early rejection of null hypotheses; efficacy stopping, or early stopping for lack of benefit; futility stopping. This way, the expected sample size required can be reduced over the more classical single-stage approach. Moreover,

multi-arm multi-stage designs, which allow multiple experimental treatments to share a control group, can increase efficiency even further (Parmar et al., 2014).

Group sequential methods are not frequently used in crossover trial settings however, in particular ones with multiple experimental treatments. Hauck et al. (1997) investigated the performance of group sequential trials for average bioequivalence employing an AB/BA crossover design, and Jennison and Turnbull (2000) provided one possible analysis method for a group sequential AB/BA crossover with a normally distributed endpoint. To the best of our knowledge, no study has explored group sequential theory for crossover trials with more than one experimental treatment being compared to a shared control.

Thus, one possible explanation for the lack of group sequential crossover trials may be that there is not yet available a formal proof of how to strongly control the familywise error rate of such a trial with multiple experimental treatments, because such a proof is usually required for regulatory approval (Wason et al., 2014). In comparison to a proof for a parallel multi-arm multi-stage design (Magirr et al., 2012), proving strong control of the familywise error rate is complicated here due to difficulties associated with the covariance structure implied by mixed model analysis. As has been remarked, multiple testing corrections for mixed models are only presently available for certain specific circumstances (Bender and Lange, 2001). Extension to this setting is particularly significant, though, given the noted advantages of comparing multiple experimental treatments to a shared control, in terms of both trial management and sample size (Parmar et al., 2014).

Potential exists, given a proof, for the efficiency of crossover trial designs to be improved. In this work, we begin by providing such a proof for a linear mixed model using period and treatment as fixed effects and individuals as random effects. Following this, using the four-treatment, four-period TOMADO trial (Quinnell et al., 2014) as an example, we explore and discuss the efficiency gains that group sequential designs could bring in a crossover setting.

2. Methods

2.1. Notation, hypotheses, and analysis

The trial is assumed to have $D \geq 2$ treatments initially, indexed $d = 0, ..., D - 1$. Treatments $d = 1, ..., D - 1$ are experimental, to be compared to the control $d = 0$. A maximum of L stages are planned for the trial. At each stage, patients are allocated to each of a set of treatment sequences, which specify an order in which a patient receives treatments. The sequences used at each stage are determined by the number of treatments remaining in the trial at that stage. Without loss of generality, we will assume that if a treatment or treatments are dropped, treatment $D - 1$ is dropped first, then $D - 2$, and so on, because treatments can always be relabeled at each interim analysis. Then, we denote by $S_r = \{s_{ri} : i = 1, ..., |S_r|\}, r = 2, ..., D$, the set of sequences for patient treatment allocation when r treatments remain in the trial, with each S_r written in the form assuming that it is exactly treatments $d = 0, ..., r - 1$ that remain. We further constrain each S_r to contain only complete block sequences that are balanced for period. Specifically, complete block allocation requires all sequences to contain each treatment remaining in the trial exactly once, and period balance requires an equal number of patients to receive each treatment remaining in the trial in each period. These

constraints allow the use of the popular Latin and Williams squares (Jones and Kenward, 2014).

A fixed group size n is used for each stage of the trial and is chosen such that at every stage each sequence is used an equal number of times. Thus, n must be divisible by the lowest common multiple of $|S_2|, ..., |S_D|$. Designing the trial in this manner ensures that each treatment is considered equally.

Outcome data are assumed to be normally distributed, and a linear mixed model is used for analysis, given by

$$y_{ijkl} = \mu_0 + \pi_j + \tau_{d[j,k,l]} + s_{ikl} + \varepsilon_{ijkl},$$

or

$$Y = X\beta + Zb + \varepsilon,$$

where

- Y is the vector of responses, containing the values of the y_{ijkl}; the response for individual i, in period j, on sequence k, in stage l,
- β is the vector of fixed effects, of length $2D - 1$, consisting of
- μ_0 the mean response on treatment 0 in period 1, an intercept term,
- π_j the fixed period effect for period j, with the identifiability constraint $\pi_1 = 0$. Note that the period is reset to 1 for each new stage of the trial. That is, the first period of stage 2 is treated as period 1 rather than period $D + 1$, also used in later stages. Thus, we have exactly $D - 1$ non-zero period effects given our restriction to complete block sequences.
- $\tau_{d[j,k,l]}$ is theτ fixed direct treatment effect for an individual in period j, on sequence k, in stage l, with the identifiability constraint $\tau_0 = 0$,
- X is the matrix linking the fixed effects to the vector of responses,
- b is the vector of random effects, consisting of the s_{ikl}; the random effect for individual i, on sequence k, in stage l,
- Z is the matrix linking the random effects to the vector of responses,
- ε is the vector of residuals, consisting of the ε_{ijkl}; the residual for individual i, in period j, on sequence k, in stage l.

Additionally, denoting by $\sigma_b^2 > 0$ and $\sigma_e^2 > 0$ the between- and within-subject variances, respectively, we take

$$\text{cov}(s_{i_1 k_1 l_1}, s_{i_2 k_2 l_2}) = \sigma_b^2 \delta_{i_1 i_2} \delta_{k_1 k_2} \delta_{l_1 l_2},$$
$$\text{cov}(\varepsilon_{i_1 j_1 k_1 l_1}, \varepsilon_{i_2 j_2 k_2 l_2}) = \sigma_e^2 \delta_{i_1 i_2} \delta_{j_1 j_2} \delta_{k_1 k_2} \delta_{l_1 l_2},$$

where δ_{ij} is the Kronecker delta function. Incorporation of fixed effects for period and treatment only and our chosen covariance structure above are the conventional choices for a crossover trial (Jones and Kenward, 2014).

We test $D - 1$ hypotheses. Because we are interested in testing the efficacy of experimental treatments in comparison to a control, we consider the case of one-sided alternative hypotheses $H_{0d} : \tau_d \leq 0, H_{1d} : \tau_d > 0$, for $d = 1, ..., D - 1$.

At each interim analysis, the above model is used to compute an estimate, $\hat{\beta}_l$ ($l = 1, ..., L$), for β through the standard maximum likelihood estimator of a linear mixed model

$$\hat{\boldsymbol{\beta}}_l = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \sim MVN\{\boldsymbol{\beta}, (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\},$$

where $\boldsymbol{\Sigma} = \boldsymbol{Z}\mathrm{cov}(\boldsymbol{b},\boldsymbol{b})\boldsymbol{Z}^{\mathrm{T}} + \mathrm{cov}(\boldsymbol{\varepsilon},\boldsymbol{\varepsilon})$ (Fitzmaurice et al., 2011). From this we acquire $\hat{\boldsymbol{\tau}}_l = (\hat{\tau}_{1l}, ..., \hat{\tau}_{D-1l})^{\mathrm{T}}$, which consists of the maximum likelihood estimates for each τ_d. Then, each $\hat{\tau}_{dl}$ is standardized to give $D-1$ test statistics $Z_{dl} = \hat{\tau}_{dl}I_{dl}^{1/2}, d = 1, ..., D-1$, with $I_{dl} = \{\mathrm{var}(\hat{\tau}_{dl})\}^{-1}$ the information level for treatment d at interim analysis l. Since $\hat{\boldsymbol{\tau}}_l$ is estimated via a normal linear model, we know that $E(Z_{dl}) = \tau_d I_{dl}^{1/2}$ (Jennison and Turnbull, 2000).

Given fixed futility boundaries, f_{dl}, and efficacy bounds, e_{dl}, the following stopping rules are used at each analysis $l = 1, ..., L$, for each experimental treatment $d = 1, ..., D-1$ satisfying $f_{dm} \leq Z_{dm} < e_{dm}$ for $m = 1, ..., l-1$

- if $Z_{dl} < f_{dl}$ treatment d is dropped without rejecting H_{0d},
- if $f_{dl} \leq Z_{dl} < e_{dl}$ the trial is continued with treatment d still present,
- and if $e_{dl} \leq Z_{dl}$ treatment d is dropped and H_{0d} rejected.

The control treatment, $d = 0$, remains present at every undertaken stage, and we only proceed to an additional stage if there is at least one experimental treatment remaining in the trial. It is convenient to take $f_{dl} = f_l$ and $e_{dl} = e_l$ for all d and l, as well as $f_L = e_L$ in order to ensure that the trial conforms to the desired maximum number of stages and so that a conclusion is made for each H_{0d}. Note that rejection of one treatment's null hypothesis does not end the trial. Furthermore, with this formulation, once a treatment is dropped from the trial, its standardized treatment effect is not tested in any future analyses.

In what follows, we will make use of the vectors $\boldsymbol{\omega}_R = (\omega_{R1}, ..., \omega_{RD-1})^T$ and $\boldsymbol{\psi}_R = (\psi_{R1}, ..., \psi_{RD-1})^T$. Here, $\omega_{Rd} \in \{1, ..., L\}$ is the analysis at which experimental treatment d was dropped from the trial. Moreover, $\psi_{Rd} \in \{0, 1\}$ with $\psi_{Rd} = 1$ if experimental treatment d was dropped for efficacy and 0 otherwise. Prior to a trial's commencement, $\boldsymbol{\omega}_R$ and $\boldsymbol{\psi}_R$ are unknown random variables. However, the probability that the trial progresses according to some particular $\boldsymbol{\omega} = (\omega_1, ..., \omega_{D-1})^T$ and $\boldsymbol{\psi} = (\psi_1, ..., \psi_{D-1})^T$, given a vector of true response rates $\boldsymbol{\tau} = (\tau_1, ..., \tau_{D-1})^T$, can be computed using multivariate normal integration. More specifically, given this particular $(\boldsymbol{\omega}, \boldsymbol{\psi})$ pair, the covariance between and the information level of the test statistics can be computed and the following integral evaluated (see Jennison and Turnbull [2000] or Wason [2015] for further details):

$$\mathrm{pr}(\boldsymbol{\omega}_R = \boldsymbol{\omega}, \boldsymbol{\psi}_R = \boldsymbol{\psi}|\boldsymbol{\tau}) = \int_{\mathrm{l}(1,\omega_1,\psi_1)}^{\mathrm{u}(1,\omega_1,\psi_1)} ... \int_{\mathrm{l}(L,\omega_{D-1},\psi_{D-1})}^{\mathrm{u}(L,\omega_{D-1},\psi_{D-1})} \phi\{\boldsymbol{x}, \boldsymbol{r}(\boldsymbol{\tau}, L) \circ \boldsymbol{I}_{(\boldsymbol{\omega},\boldsymbol{\psi})}^{1/2}, \Lambda_{(\boldsymbol{\omega},\boldsymbol{\psi})}\}\mathrm{d}x_{L(D-1)} ... \mathrm{d}x_{11},$$

where

- $\boldsymbol{x} = (x_{11}, ..., x_{1(D-1)}, ..., x_{L1}, ..., x_{L(D-1)})^T$,
- $\phi\{\boldsymbol{x}, \boldsymbol{\mu}, \Lambda\}$ is the probability density function of a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Λ, evaluated at vector $\boldsymbol{x}$,
- $\boldsymbol{r}(\boldsymbol{\tau}, L)$ is the vector formed by repeating $\boldsymbol{\tau}$ L times,

- $\boldsymbol{I}_{(\omega,\psi)} = \left(\boldsymbol{I}_{1,(\omega,\psi)}^{\mathrm{T}}, \ldots, \boldsymbol{I}_{L,(\omega,\psi)}^{\mathrm{T}}\right)^{\mathrm{T}}$, where $\boldsymbol{I}_{l,(\omega,\psi)} = (I_{1l}, \ldots, I_{(D-1)l})_{(\omega,\psi)}^{T}$ is the vector of information
  levels for the estimated treatment effects at interim analysis l, according to (conditional on) the particular (ω, ψ) being considered,
- $\circ$ denotes the Hadamard product of two vectors,
- the square root of the vector $\boldsymbol{I}_{(\omega,\psi)}$ is taken in an element-wise manner,
- l and u are functions that tell us the lower and upper integration limits for the test statistic Z_{dl} given values for l, ω_d and ψ_d. For example, $l(1,2,1) = f_1$ and $u(1,2,1) = e_1$, while $l(2,2,1) = e_2$ and $u(2,2,1) = \infty$, and then $l(l,2,1) = -\infty$ and $u(l,2,1) = \infty$ for $l > 2$,
- $\Lambda_{(\omega,\psi)}$ is the covariance matrix between the standardized test statistics at and across each interim analysis according to (ω, ψ). Thus, using $\boldsymbol{Z}_l = (Z_{1l}, \ldots, Z_{D-1l})^{\mathrm{T}}$, we have

$$\Lambda(\omega, \psi) = \begin{pmatrix} \mathrm{cov}(\boldsymbol{Z}_1, \boldsymbol{Z}_1 | \omega, \psi) & \ldots & \mathrm{cov}(\boldsymbol{Z}_1, \boldsymbol{Z}_L | \omega, \psi) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(\boldsymbol{Z}_L, \boldsymbol{Z}_1 | \omega, \psi) & \ldots & \mathrm{cov}(\boldsymbol{Z}_L, \boldsymbol{Z}_L | \omega, \psi) \end{pmatrix}.$$

However, $\boldsymbol{Z}_l = \hat{\boldsymbol{\tau}}_l \circ \boldsymbol{I}_{l,(\omega,\psi)}^{1/2}$, and by the properties of normal linear models $\mathrm{cov}(\hat{\boldsymbol{\tau}}_{l_1}, \hat{\boldsymbol{\tau}}_{l_2} | \omega, \psi) = \mathrm{cov}(\hat{\boldsymbol{\tau}}_{l_2}, \hat{\boldsymbol{\tau}}_{l_2} | \omega, \psi)$ $(l_1, l_2 = 1, \ldots, L; l_1 \leq l_2)$ (Jennison and Turnbull, 2000), giving

$$\mathrm{cov}(\boldsymbol{Z}_{l_1}, \boldsymbol{Z}_{l_2} | \omega, \psi) = \mathrm{diag}(\boldsymbol{I}_{l_1,(\omega,\psi)}^{1/2})\mathrm{cov}(\hat{\boldsymbol{\tau}}_{l_2}, \hat{\boldsymbol{\tau}}_{l_2} | \omega, \psi)\mathrm{diag}(\boldsymbol{I}_{l_2,(\omega,\psi)}^{1/2}), \tag{2.1}$$

for $l_1, l_2 = 1, \ldots, L, l_1 \leq l_2$, and where $\mathrm{diag}(v)$ is the matrix formed by placing the elements of vector v along the leading diagonal.

Note that equation (2.1) in conjunction with the expectations of our standardized test statistics and the observation that $(\boldsymbol{Z}_1^{\mathrm{T}}, \ldots, \boldsymbol{Z}_L^{\mathrm{T}})^{\mathrm{T}}$ is multivariate normal can be restated simply as that our test statistics follow the canonical joint distribution (Jennison and Turnbull, 2000).

2.2. Familywise error rate control

It is a common requirement of clinical trial designs that the probability of one or more false rejections within the family of null hypotheses is not greater than some α. This is known as strong control of the familywise error rate. In this section, we establish strong control for our considered trial design.

To evaluate the familywise error rate of a design, for any τ, the above integral can be evaluated for all ω and ψ that would imply that a type-I error is made and the results summed. In order to demonstrate how to strongly control, though, it is essential to know the forms of the $\boldsymbol{I}_{l,(\omega,\psi)}$ and $\Lambda_{(\omega,\psi)}$ for each (ω, ψ). However, by equation (2.1), the $\boldsymbol{I}_{l,(\omega,\psi)}$ and $\Lambda_{(\omega,\psi)}$ can be determined if $\mathrm{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | \omega, \psi)$ is known for $l = 1, \ldots, L$.

Thus, consider the matrix $\mathrm{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | \omega, \psi)$ for some $l \leq L$ and any (ω, ψ). We compute values for L_{lr} $(r = 1, \ldots, D)$, the number of stages of the trial, up to analysis l, in which r treatments were remaining. Because we do not continue the trial unless at least one experimental treatment remains, $L_{l1} = 0$ always. It will be convenient, however, to still include this value. Moreover, it is clear that the L_{lr} are uniquely determined given

$(\boldsymbol{\omega}, \boldsymbol{\psi})$. Now, $\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | \boldsymbol{\omega}, \boldsymbol{\psi})$ can always be decomposed to be a sum over the determined L_{lr} and the pre specified sequences S_r (see Fitzmaurice et al. [2011] for details):

$$\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | \boldsymbol{\omega}, \boldsymbol{\psi}) = \text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD}), = \left(\sum_{r=1}^{D} L_{lr} \frac{n}{|S_r|} \sum_{i=1}^{|S_r|} \boldsymbol{X}_{s_{ri}}^{\mathrm{T}} \boldsymbol{\Sigma}_r^{-1} \boldsymbol{X}_{s_{ri}} \right) .$$

Here $\boldsymbol{X}_{s_{ri}}$ is the uniquely defined $r \times (2D - 1)$ design matrix for a single patient allocated to sequence s_{ri}, and $\boldsymbol{\Sigma}_r$ is the easily computed $r \times r$ covariance matrix of the responses for a single patient allocated r treatments in total. The factor $n/|S_r|$ arises from the number of patients allocated to each sequence s_{ri} by our choice of period balance.

We now establish two key results about $\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD})$. Following this, we provide a proof detailing how to strongly control the familywise error rate.

Theorem 2.1. *Let $\boldsymbol{\beta} = (\mu_0, \pi_2, ..., \pi_D, \tau_1, ..., \tau_{D-1})^T$. Consider an analysis to be performed after some number of stages l. Then*

1. *We have*

$$\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD-1} = 0, L_{lD} = l) = \left(\frac{ln}{|S_D|} \sum_{i=1}^{|S_D|} \boldsymbol{X}_{s_{Di}}^{\mathrm{T}} \boldsymbol{\Sigma}_D^{-1} \boldsymbol{X}_{s_{Di}} \right)^{-1},$$

$$= \frac{1}{ln} \begin{pmatrix} F & \boldsymbol{G}^{\mathrm{T}} & \boldsymbol{G}^{\mathrm{T}} \\ \boldsymbol{G} & \boldsymbol{H} & \boldsymbol{0}_{D-1,D-1} \\ \boldsymbol{G} & \boldsymbol{0}_{D-1,D-1} & \boldsymbol{H} \end{pmatrix}, \qquad (2.2)$$

    where

$$F = \sigma_b^2 + \frac{2D-1}{D}\sigma_e^2,$$
$$\boldsymbol{G}_{pq} = -\sigma_e^2 \qquad (p = 1, ..., D-1; q = 1),$$
$$\boldsymbol{H}_{pq} = \sigma_e^2(1 + \delta_{pq}) \qquad (p = 1, ..., D-1; q = 1, ..., D-1).$$

2. *If $q \geq 2$ is the largest integer such that $L_{lr} = 0$ for $r = 1, ..., q - 1$, then the covariance of the estimates of the fixed effects $\hat{\pi}_{2l}, ..., \hat{\pi}_{ql}, \hat{\tau}_{1l}, ..., \hat{\tau}_{q-1l}$ is identical to what it would be for $L_{l1} = \cdots = L_{lD-1} = 0$. Moreover, the covariance between the estimates of $\hat{\pi}_{2l}, ..., \hat{\pi}_{ql}, \hat{\tau}_{1l}, ..., \hat{\tau}_{q-1l}$ and the estimates of $\hat{\pi}_{q+1l}, ..., \hat{\pi}_{Dl}, \hat{\tau}_{ql}, ..., \hat{\tau}_{D-1l}$ is also identical to what it would be for $L_{l1} = \cdots = L_{lD-1} = 0$.*

Proof. See Appendix C.

Note that part (1) of the above theorem implies

$$\text{cov}(\hat{\tau}_{d_1l}, \hat{\tau}_{d_2l} | L_{l1} = \cdots = L_{lD-1} = 0, L_{lD} = l) = \frac{\sigma_e^2}{ln}(1 + \delta_{d_1d_2}),$$

for $d_1, d_2 \in \{1, ..., D - 1\}$. This is the familiar result for complete block sequences that there is no dependence upon the between-patient variance σ_b^2 (Jones and Kenward, 2014). $\square$

Theorem 2.2. *A group sequential crossover trial of the type considered, with $D \geq 2$, testing the $D - 1$ hypotheses $H_{0d} : \tau_d \leq 0, H_{1d} : \tau_d > 0$, attains a maximal value of its familywise error rate for $\tau_1 = \cdots = \tau_{D-1} = 0$.*

Proof. Theorem 2.1 implies that elements of the covariance matrix $\mathrm{cov}(\hat{\tau}_l, \hat{\tau}_l)$ that differ from the case where no treatments have been dropped are exactly those corresponding to unstandardized test statistics no longer of importance. Consequently, the values of $I_{l,(\omega,\psi)}$ and $\Lambda_{(\omega,\psi)}$ that differ from the case $\omega = (L, ..., L)^{\mathrm{T}}$ are only ever those corresponding to limits of integration given by $(-\infty, \infty)$ in our computation of $\mathrm{pr}(\omega_R = \omega, \psi_R = \psi | \tau)$. By the marginal distribution properties of the multivariate normal distribution, we therefore need only consider one matrix $\Lambda_{(\omega,\psi)}$ and one set of vectors $I_{l,(\omega,\psi)}$ ($l = 1, ..., L$), exactly those given by the case $\omega = (L, ..., L)^{\mathrm{T}}$. Denote these by Λ and I_l, and set $I = (I_1^{\mathrm{T}}, ..., I_L^{\mathrm{T}})^{\mathrm{T}}$. For more information on this, see Appendix A. We now have

$$\mathrm{pr}(\omega_R = \omega, \psi_R = \psi | \tau) = \int_{\mathrm{l}(1,\omega_1,\psi_1)}^{\mathrm{u}(1,\omega_1,\psi_1)} ... \int_{\mathrm{l}(L,\omega_{D-1},\psi_{D-1})}^{\mathrm{u}(L,\omega_{D-1},\psi_{D-1})} \phi\left\{x, r(\tau, L) \circ I^{1/2}, \Lambda\right\} dx_{L(D-1)}...dx_{11}.$$

Now, consider without loss of generality the probability that we reject H_{01}, and denote by Ω and Ψ the sets of all possible ω and ψ, respectively. By integrating over all possible values of $\omega_2, ..., \omega_{D-1}$ and $\psi_2, ..., \psi_{D-1}$, we have that the probability that we reject each H_{01} does not depend on the values of $\tau_2, ..., \tau_{D-1}$; that is, on the other treatments tested:

$$\mathrm{pr}(\text{Reject } H_{01} | \tau) = \sum_{\{\psi \in \Psi : \psi_1 = 1\}} \sum_{\omega \in \Omega} \mathrm{pr}(\omega_R = \omega, \psi_R = \psi | \tau),$$

$$= \sum_{\{\psi \in \Psi : \psi_1 = 1\}} \sum_{\omega \in \Omega} \int_{\mathrm{l}(1,\omega_1,\psi_1)}^{\mathrm{u}(1,\omega_1,\psi_1)} ... \int_{\mathrm{l}(L,\omega_{D-1},\psi_{D-1})}^{\mathrm{u}(L,\omega_{D-1},\psi_{D-1})} \phi\{x, r(\tau, L) \circ I^{1/2}, \Lambda\}$$

$$dx_{L(D-1)}...dx_{11},$$

$$= \sum_{\omega_1 = 1}^{L} \int_{\mathrm{l}(1,\omega_1,1)}^{\mathrm{u}(1,\omega_1,1)} ... \int_{\mathrm{l}(L,\omega_1,1)}^{\mathrm{u}(L,\omega_1,1)} \phi\{(x_{11}, ..., x_{L1})^{\mathrm{T}}, r(\tau_1, L) \circ I_{\tau_1}^{1/2}, \Lambda_{\tau_1}\}$$

$$dx_{L1}...dx_{11},$$

where I_{τ_1} and Λ_{τ_1} are the restrictions of I and Λ to rows and columns corresponding to experimental treatment $d = 1$ respectively. This final form for $\mathrm{pr}(\text{Reject } H_{01} | \tau)$ is identical to what it would be in the case $D = 2$. Therefore, to ascertain the τ giving the maximal familywise error rate of a trial with $D \geq 2$, it suffices to consider which $\tau_* \leq 0$ maximizes the probability that H_{01} is rejected in a trial with $D = 2$ initial treatments. For then, $\tau = (\tau_*, ..., \tau_*)^{\mathrm{T}}$ using this τ_* will provide the maximum probability of rejecting at least one true H_{0d} for some d; that is, the maximum familywise error rate. To see this, consider the familywise error rate for $\tau = (\tau_*, ..., \tau_*)^{\mathrm{T}}$. If one changes some individual element τ_{d_1} of this vector, this does not effect the probability that H_{0d_2} is rejected for $d_2 \neq d_1$, and it can only decrease the probability that H_{0d_1} is incorrectly rejected. Thus, overall, straying from this $\tau = (\tau_*, ..., \tau_*)^{\mathrm{T}}$ can only decrease the familywise error rate.

Thus, now consider all possible realizations of the test statistics of a trial with $D = 2$ and their associated values of $(\omega, \psi) = (\omega_1, \psi_1)$. We have $Z = (Z_{11}, ..., Z_{1L})^{\mathrm{T}} \in \mathbb{R}^L$, with $Z_{1L} = \cdots = Z_{1\omega_1}$ if the trial was stopped at stage ω_1. Now consider increasing the value of the test statistics by some $\eta > 0$. All instances before where H_{01} was rejected will still exceed the efficacy bound of that stage, or earlier, and so H_{01} will still be rejected. Therefore, the probability of rejecting H_{01} is at least as large as before. Thus, increasing

the value of $\tau_1 \leq 0$ causes a non-decreasing change in the value of the type-I error rate. Therefore, the probability of rejecting H_{01} is maximized by $\tau_1 = 0$, implying in turn that the maximal familywise error rate of a trial with $D \geq 2$ is given by $\boldsymbol{\tau} = (\tau_1, ..., \tau_{D-1})^{\mathrm{T}} = (0, ..., 0)^{\mathrm{T}}$. $\qquad\square$

2.3. Design characteristics

A trial will now be fully specified given values for D, L, σ_e^2, and n, as well as choices for S_r and the futility and efficacy boundaries $f_1, ..., f_L$ and $e_1, ..., e_L$, respectively. Given these, Λ and $\boldsymbol{I}$ can be computed using the results above. Then, by Theorem 2.2 we can strongly control the familywise error rate to α for this design using the following sum of integrals:

$$\alpha = \sum_{\{\boldsymbol{\psi} \in \Psi : \Sigma_d \psi_d > 0\}} \sum_{\boldsymbol{\omega} \in \Omega} \int_{\mathrm{l}(1,\omega_1,\psi_1)}^{\mathrm{u}(1,\omega_1,\psi_1)} ... \int_{\mathrm{l}(L,\omega_{D-1},\psi_{D-1})}^{\mathrm{u}(L,\omega_{D-1},\psi_{D-1})} \phi\{\boldsymbol{x}, \boldsymbol{r}(0, L(D-1)), \Lambda\} dx_{L(D-1)}...dx_{11}.$$

Additionally, suppose that we wish to power this trial to reject a particular null hypothesis, without loss of generality H_{01}, at some clinically relevant difference $\tau_1 = \delta$. The type-II error rate β for H_{11} is then given by

$$\beta = 1 - \sum_{\omega_1=1}^{L} \int_{\mathrm{l}(1,\omega_1,1)}^{\mathrm{u}(1,\omega_1,1)} ... \int_{\mathrm{l}(L,\omega_1,1)}^{\mathrm{u}(L,\omega_1,1)} \phi\left\{(x_{11}, ..., x_{L1})^{\mathrm{T}}, \boldsymbol{r}(\delta, L) \bigcirc \boldsymbol{I}_{\tau_1}^{1/2}, \Lambda_{\tau_1}\right\} dx_{L1}...dx_{11}.$$

Moreover, denoting by N and O the total number of patients and observations required by the trial, respectively, we can compute the expected sample size, $E(N|\boldsymbol{\tau})$, or expected number of observations, $E(O|\boldsymbol{\tau})$, for any $\boldsymbol{\tau}$, according to

$$E(N|\boldsymbol{\tau}) = \sum_{\boldsymbol{\psi} \in \Psi} \sum_{\boldsymbol{\omega} \in \Omega} \mathrm{pr}(\boldsymbol{\omega}_R = \boldsymbol{\omega}, \boldsymbol{\psi}_R = \boldsymbol{\psi}|\boldsymbol{\tau}) \mathrm{N}(\boldsymbol{\omega}, \boldsymbol{\psi}),$$

$$E(O|\boldsymbol{\tau}) = \sum_{\boldsymbol{\psi} \in \Psi} \sum_{\boldsymbol{\omega} \in \Omega} \mathrm{pr}(\boldsymbol{\omega}_R = \boldsymbol{\omega}, \boldsymbol{\psi}_R = \boldsymbol{\psi}|\boldsymbol{\tau}) \mathrm{O}(\boldsymbol{\omega}, \boldsymbol{\psi}).$$

Here, $\mathrm{N}(\boldsymbol{\omega}, \boldsymbol{\psi})$ and $\mathrm{O}(\boldsymbol{\omega}, \boldsymbol{\psi})$ are functions that give the number of patients and observations, respectively, required by a trial that progresses according to $(\boldsymbol{\omega}, \boldsymbol{\psi})$. Specifically

$$\mathrm{N}(\boldsymbol{\omega}, \boldsymbol{\psi}) = n \max_{\{d=1,...,D-1\}} \omega_d,$$

$$\mathrm{O}(\boldsymbol{\omega}, \boldsymbol{\psi}) = n \sum_{l=1}^{L} \left(\sum_{d=1}^{D-1} \mathbb{I}_{\{\omega_d \geq l\}} + 1 \right),$$

where $\mathbb{I}_{\{\omega_d \geq l\}} = 1$ if $\omega_d \geq l$ and 0 otherwise.

3. Example: TOMADO

As an example of how to design a group sequential crossover trial with strong control of the familywise error rate, we will make use of the TOMADO crossover randomized controlled trial (Quinnell et al., 2014). This open-label trial compared three experimental treatments to a single control for the treatment of sleep apnea-hypopnea using a four-treatment four-period crossover design. The normally distributed secondary endpoint,

Epworth Sleepiness Scale, was used to observe negative test statistics. Therefore, we consider the decrease as the endpoint in order to retain the same hypothesis tests as before $H_{0d} : \tau_d \leq 0, H_{1d} : \tau_d > 0$, $d = 1, 2, 3$. The trial planned to recruit 90 patients, and utilizing restricted error maximum likelihood estimation, the final analysis calculated that $\sigma_e^2 = 6.51$. Taking this variance as the truth, the trial had a familywise error rate $\alpha = 0.05$ for $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)^{\mathrm{T}} = (0, 0, 0)^{\mathrm{T}}$, and $\beta = 0.2$ for H_{11} at $\tau_1 = 1.11$.

Many methods exist for determining boundaries for a one-sided group sequential trial with parallel treatment arms. Here, we consider analogues of the power family boundaries of Pampallona and Tsiatis (1994). For this, values for the desired type-I and type-II error rates, a clinically relevant difference δ, the maximum number of stages L, the within-person variance σ_e^2, and a shape parameter Δ must be specified. A two-dimensional grid search is then used to find the exact required maximal sample size. From this, a suitable value of n is identified by rounding up to the nearest integer such that n is as required divisible by $|S_2|, ..., |S_D|$. Utilizing Williams squares for our designs, n was forced to be divisible by 12.

Taking $\alpha = 0.05, \beta = 0.2, \delta = 1.11, \sigma_e^2 = 6.51$, $L = 3$, and $\Delta = -0.25, 0, 0.5, 0.5$ as examples, group sequential crossover trial designs were determined and compared to the single-stage design used by TOMADO. All computations were done in R (R Core Team, 2016) using the package groupSeqCrossover, available from https://github.com/mjg211/article_code. Matlab (The Mathworks Inc., 2016) code employing symbolic algebra is also available to return the matrices given by several of the equations in the text. Use of both the R and Matlab code is detailed in Appendix D.

A summary of the performance of the designs is provided in Table 1, and their computed boundaries are displayed in Figure 1. We can see that, as is the case for two-arm parallel trial designs, there is a trend that larger values of Δ result in larger maximum sample sizes and lower expected sample sizes due to their larger stopping regions. However, this is not the case for $\Delta = 0.25$ because of the requirement to round to a suitable integer value of n.

Plots of the probability of rejecting H_{01} and rejecting H_{0d} for some $d = 1, 2, 3$ are provided for a range of values of θ when $\boldsymbol{\tau} = (\theta, \theta, \theta)^{\mathrm{T}}$ in Figure 2. The power curves are similar for all of the designs, with the only differences a result of rounding in the group sequential designs to achieve suitable values of n.

Table 1. Example design performance. Summary of the performance of the single-stage and considered group sequential designs.[a]

	Design				
	Single-stage	$\Delta = -0.25$	$\Delta = 0$	$\Delta = 0.25$	$\Delta = 0.5$
n	90	36	36	48	48
pr(Reject $H_{01}|\boldsymbol{\tau} = \boldsymbol{0}$)	0.02	0.02	0.02	0.02	0.02
pr(Reject $H_{01}|\boldsymbol{\tau} = \boldsymbol{\delta}$)	0.80	0.85	0.83	0.90	0.83
pr(Reject H_{0d} for some $d|\boldsymbol{\tau} = \boldsymbol{0}$)	0.05	0.05	0.05	0.05	0.05
pr(Reject H_{0d} for some $d|\boldsymbol{\tau} = \boldsymbol{\delta}$)	0.95	0.97	0.97	0.98	0.97
$E(N|\boldsymbol{\tau} = \boldsymbol{0})$	90.0	76.8	70.0	82.6	69.6
$E(N|\boldsymbol{\tau} = \boldsymbol{\delta})$	90.0	100.3	95.7	110.7	98.9
$E(O|\boldsymbol{\tau} = \boldsymbol{0})$	360.0	269.3	240.3	283.1	244.5
$E(O|\boldsymbol{\tau} = \boldsymbol{\delta})$	360.0	367.2	341.8	380.4	327.7
maxN	90	108	108	144	144
maxO	360	432	432	576	576

[a]The number of decimal places displayed in each row indicates the number to which rounding was performed.
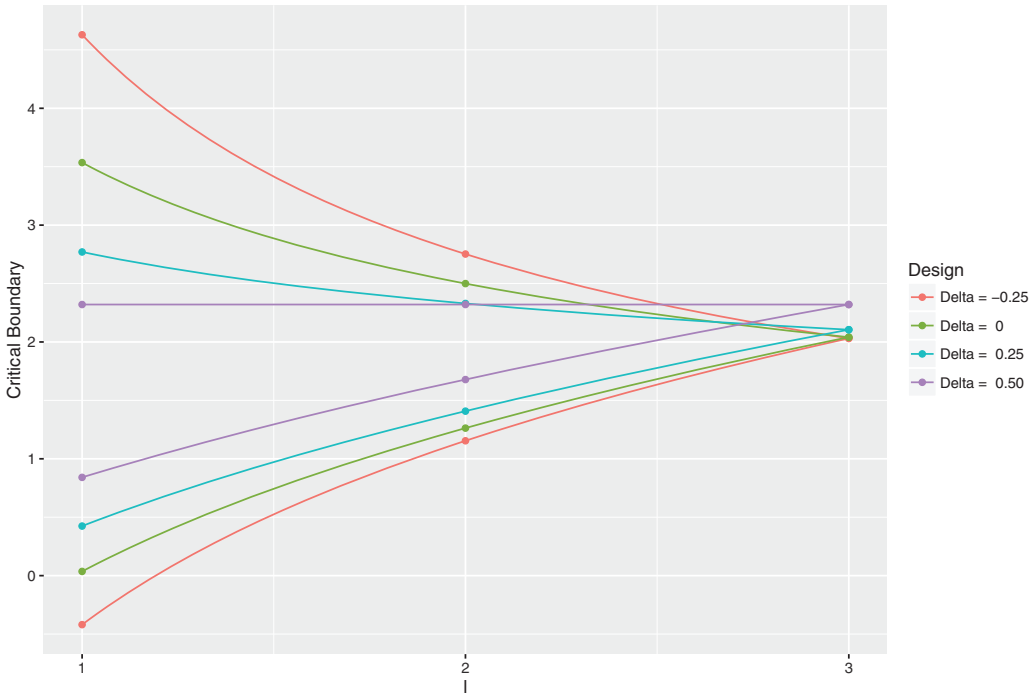
Figure 1. Stopping boundaries. Computed efficacy and futility boundaries of the considered group sequential designs.

As is to be expected for group sequential designs, the maximum sample size and maximum number of observations are larger than those for the single-stage design. However, the group sequential designs have lower expected sample sizes under the global null hypothesis ($\tau = \mathbf{0} = (0, 0, 0)^{\mathrm{T}}$), up to a maximum of 23% for $\Delta = 0.5$. This does, however, come at the expense of an increased expected sample size under the global alternative hypothesis ($\tau = \boldsymbol{\delta} = (\delta, \delta, \delta)^{\mathrm{T}}$).

From Figure 3, the expected sample sizes of the group sequential designs can be seen to be far lower than those of the single-stage design for more extreme values of θ. A similar statement holds for the expected number of observations. However, in this instance for $\Delta = 0$, 0.5, the performance of the group sequential designs is better than the single-stage design across all values of θ.

4. Discussion

There is a long history on group sequential clinical trials. Very few, however, utilize a crossover design. This may be at least in part due to no formal proof existing for how to strongly control the familywise error rate of such a trial. Here, we provided such a proof and then explored the performance of several sequential designs for the TOMADO trial.

The expected sample size of the sequential designs was observed to be far lower than that of the single-stage design for a large range of values of the true response rate on all
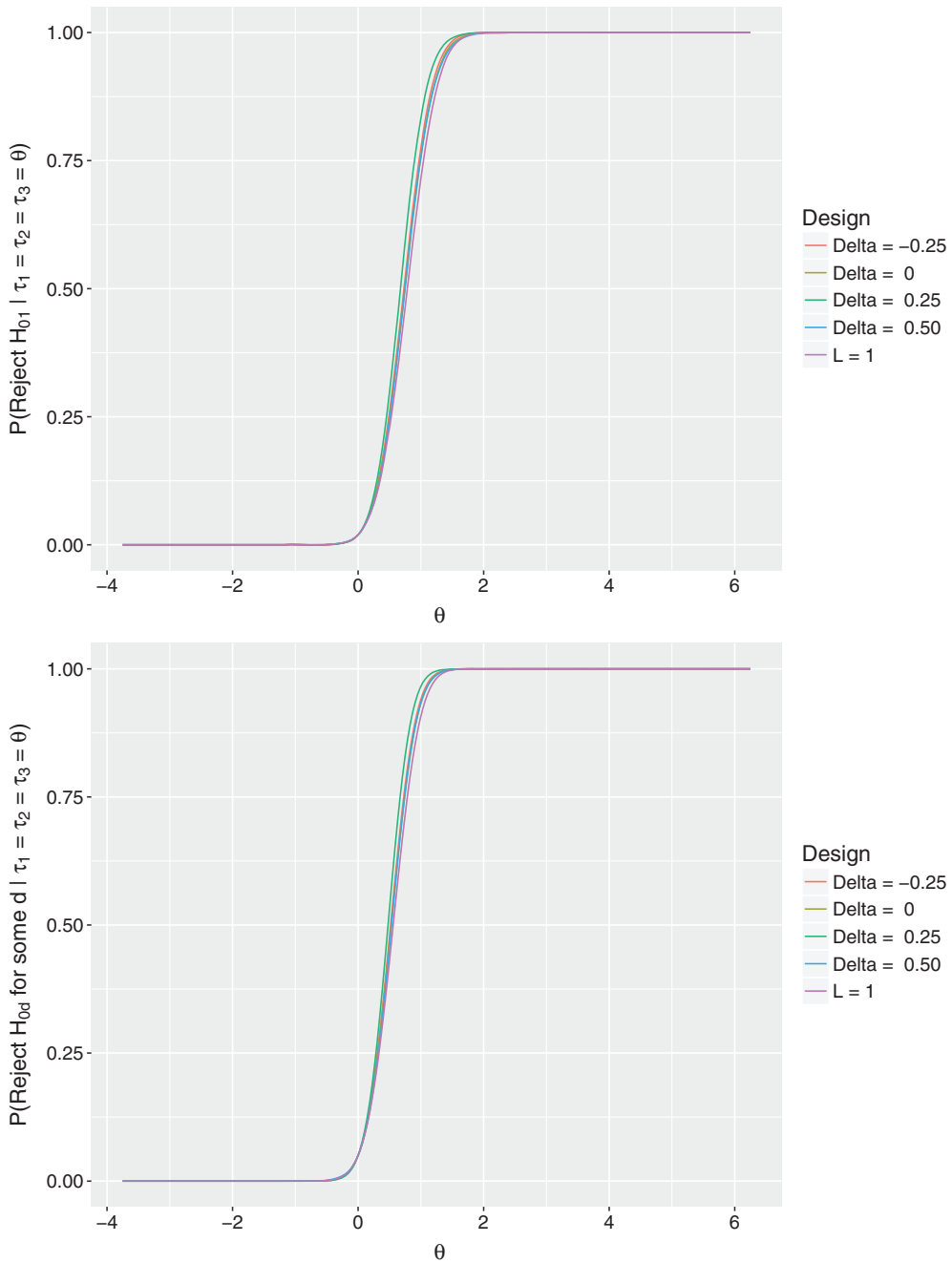
Figure 2. Power curves. Power curves of the single-stage ($L = 1$) and considered group sequential designs across a range of values of the true response rate in the experimental treatment arms θ.

experimental treatments. Unfortunately, but unsurprisingly given that the trial is not stopped unless all experimental treatments are dropped, there are regions in which the sequential designs are less efficient. Indeed, this region includes some values of θ
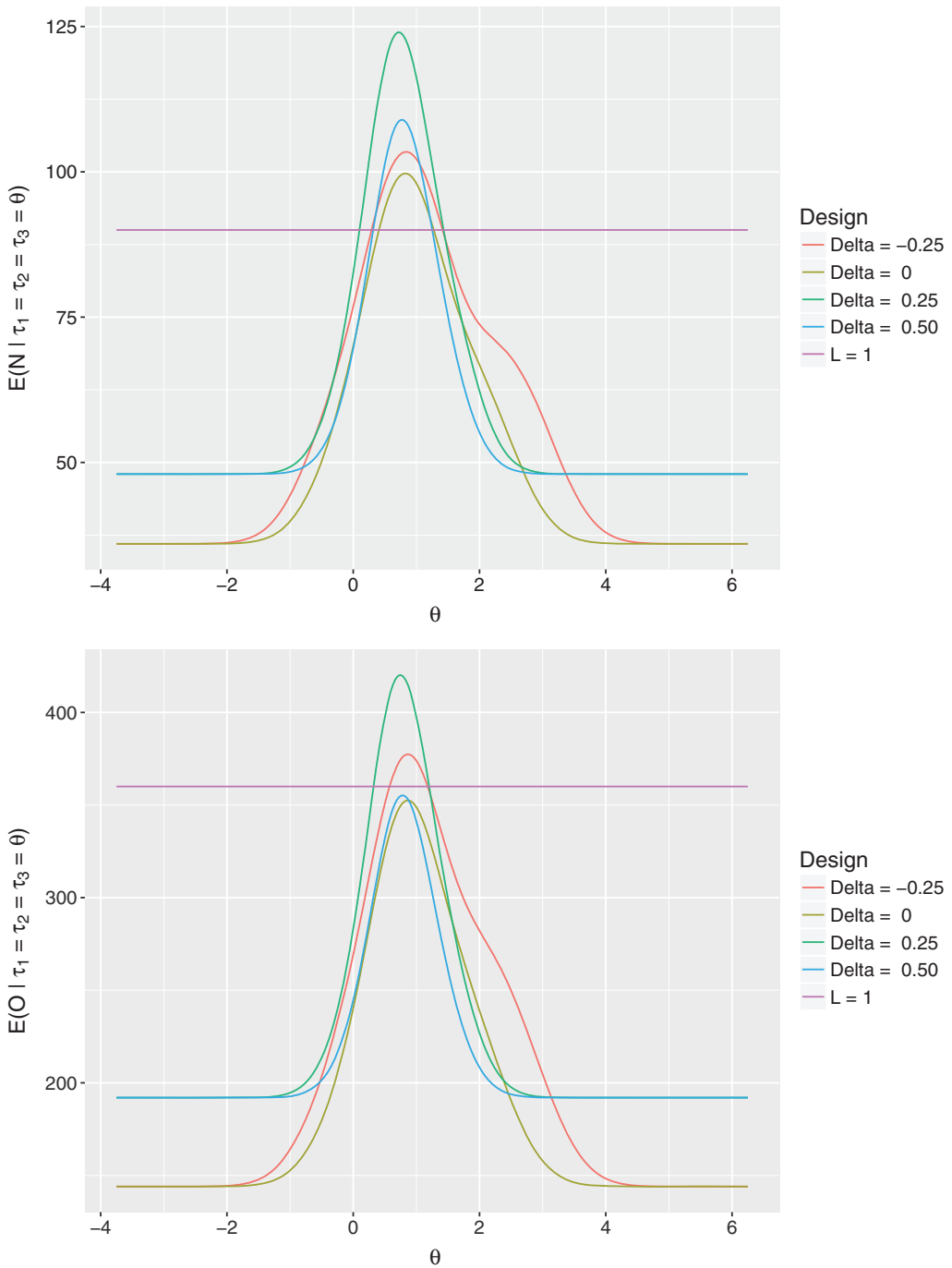
Figure 3 Performance measurement curves. Curves of the expected sample size and expected number of observations of the single-stage ($L = 1$) and considered group sequential designs across a range of values of the true response rate in the experimental treatment arms θ.

between 0 and δ, which may be more realistic observed treatment effects. However, for some considered designs, this region is very small and does not include values near 0, which is notable for ethical reasons. This issue could even be further alleviated by utilizing optimal stopping boundaries, as has been proposed for parallel arm designs (Wason and Jaki, 2012; Wason et al., 2012). Importantly, several of the designs always performed better than the single-stage design in terms of the expected number of observations required, which could be a significant factor in the cost and length of a trial. Consequently, we can conclude that a group sequential approach to a crossover trial improves efficiency in some circumstances.

Several possible extensions to our work present themselves. For example, we assumed that the period was reset in each trial stage. This could reflect a scenario where it is believed being enrolled in the trial will alter a patient's behavior. However, in some cases, such as to deal with seasonal effects, it would be preferential to have different period effects in each stage.

One simple extension would be to employ non-inferiority tests, from our present superiority testing framework. Non-inferiority tests, seeking to determine whether new treatments are not clinically worse than an established control, would have hypotheses shifted by some factor from the ones presented here. Theorem 2.2 could easily be altered to accommodate this and then popular methods for boundary determination in this setting could be applied.

Here, we have worked under an idealized scenario, assuming the within-patient variance to be known prior to trial commencement. Though this is a common assumption in group sequential theory, it does have limitations, because often a good estimate for the key variance parameter cannot be provided at the design stage. In this instance, group sequential t-tests would almost certainly be required. Furthermore, simulation is required to quantify error rates accurately in the case of small sample sizes. To explore this scenario, we analyzed the true familywise error rate under the global null hypothesis of a particular design motivated again by the TOMADO trial but with $L = 2$ and $n = 12$. We found that provided that restricted error maximum likelihood was utilized, there was very little inflation in the familywise error rate over the nominal level α. Details of this are provided in Appendix B.

Moreover, we have only explored designing group sequential crossover trials. It is well known that if a final analysis is performed on data acquired in a sequential trial, not taking into account the sequential nature, then biased treatment effects will be acquired. Extending established methodology for parameter estimation to our scenario will thus be important.

Finally, we have implicitly assumed that there will be no patient dropout and have not discussed the issue of patient recruitment rates. Though these are problems for all adaptive designs, it is important to give them note. Due to our need for one stages, data to be analyzed before the commencement of the following stage, it is likely that the length of a trial using our approach would be longer for certain recruitment rates. It could be that recruitment is paused at interim or that patients are continually

recruited under the old scheme until results are available, which would lead to overrun and an increase in the expected number of observations and sample size. Thus, this would be an important factor to consider when choosing an appropriate design for a trial.

Nevertheless, for future crossover trials, consideration should be given to a group sequential approach. This may substantially assist in the efficient prioritization of efficacious treatments.

Appendix A: Further technical details

As discussed in Section 2, part (1) of Theorem 2.1 implies that

$$\operatorname{cov}(\hat{\tau}_{d_1 l}, \hat{\tau}_{d_2 l} | L_{l1}, ..., L_{lD-1} = 0, L_{lD} = l) = \frac{\sigma_e^2}{ln}(1 + \delta_{d_1 d_2}),$$

for $d_1, d_2 \in \{1, ..., D-1\}$. Alternatively, it tells us that in this case $I_{dl} = ln/(2\sigma_e^2)$, for $d = 1, ..., D-1$.

Moreover, using the above along with equation (2.1), in conjunction with part (2) of Theorem 2.1, we have that if $f_p \leq Z_{d_1 p} < e_p$ for $p = 0, ..., l_1 - 1$ (i.e., if treatment d_1 is present up to stage l_1) and $f_q \leq Z_{d_2 q} < e_q$ for $q = 0, ..., l_2 - 1$ (i.e., if treatment d_2 is present up to stage l_2), with $l_1 \leq l_2, l_1, l_2 \in \{1, ..., L\}$, and $d_1, d_2 \in \{1, ..., D-1\}$ (taking $f_0 = -\infty$ and $e_0 = \infty$),

$$I_{d_1 a} = \frac{an}{2\sigma_e^2},$$

$$I_{d_2 b} = \frac{bn}{2\sigma_e^2},$$

$$\operatorname{cov}(Z_{d_1 a}, Z_{d_2 b} | \boldsymbol{\omega}, \boldsymbol{\psi}) = I_{d_1 a}^{1/2} \operatorname{cov}(\hat{\tau}_{d_1 b}, \hat{\tau}_{d_2 b} | L_{b1}, ..., L_{bD}) I_{d_2 b}^{1/2}$$

$$= \left(\frac{an}{2\sigma_e^2}\right)^{1/2} \frac{\sigma_e^2}{bn}(1 + \delta_{d_1 d_2})\left(\frac{bn}{2\sigma_e^2}\right)^{1/2},$$

$$= \frac{1}{2}\left(\frac{a}{b}\right)^{1/2}(1 + \delta_{d_1 d_2}),$$

for $a \leq b, a = 1, ..., l_1$, and $b = 1, ..., l_2$.

For further clarity, as an example, consider the case $D = 3$, $L = 2$, and the associated value of $\operatorname{pr}(\boldsymbol{\omega}_R = \boldsymbol{\omega}, \boldsymbol{\psi}_R = \boldsymbol{\psi} | \boldsymbol{\tau})$ when $\boldsymbol{\omega} = (2, 1)^{\mathrm{T}}$ and $\boldsymbol{\psi} = (1, 0)^{\mathrm{T}}$. Using the above, we know the following elements of the matrix $\Lambda_{(\omega,\psi)}$ and vector $\boldsymbol{I}_{(\omega,\psi)} = (\boldsymbol{I}_{1,(\omega,\psi)}^{\mathrm{T}}, \boldsymbol{I}_{2,(\omega,\psi)}^{\mathrm{T}})^{\mathrm{T}}$:

Table B.1. Performance of the small sample size group sequential crossover trial design under four analysis procedures. Specifically, pr(reject H_{0d} for some $d | \tau = 0$) is shown for each procedure to three decimal places based on 10,000 trial simulations.[a]

Procedure	Estimation	Boundary adjustment	pr(reject H_{0d} for some $d|\tau = 0$)
Procedure 1	ML	No	0.077
Procedure 2	ML	Yes	0.062
Procedure 3	REML	No	0.055
Procedure 4	REML	Yes	0.051

[a]ML = Maximum likelihood, REML = restricted error maximum likelihood.

$$\Lambda_{(\boldsymbol{\omega},\boldsymbol{\psi})} = \begin{pmatrix} 1 & \frac{1}{2} & \left(\frac{1}{2}\right)^{1/2} & \bullet \\ \frac{1}{2} & 1 & \frac{1}{2}\left(\frac{1}{2}\right)^{1/2} & \bullet \\ \left(\frac{1}{2}\right)^{1/2} & \frac{1}{2}\left(\frac{1}{2}\right)^{1/2} & 1 & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{pmatrix},$$

$$\boldsymbol{I}_{(\boldsymbol{\omega},\boldsymbol{\psi})} = \left(\frac{n}{2\sigma_e^2}, \frac{n}{2\sigma_e^2}, \frac{2n}{2\sigma_e^2}, \bullet\right)^{\mathrm{T}},$$

where we have used $\bullet$ to signify an element that we do not know the value of.

Now consider our computation of $\mathrm{pr}(\boldsymbol{\omega}_R = \boldsymbol{\omega}, \boldsymbol{\psi}_R = \boldsymbol{\psi}|\boldsymbol{\tau})$. We have

$$\mathrm{pr}(\boldsymbol{\omega}_R = \boldsymbol{\omega}, \boldsymbol{\psi}_R = \boldsymbol{\psi}|\boldsymbol{\tau}) = \int_{f_1}^{e_1} \int_{-\infty}^{f_1} \int_{e_2}^{\infty} \int_{-\infty}^{\infty} \phi\{\boldsymbol{x}, \boldsymbol{r}(\boldsymbol{\tau}, 2)\bigcirc\boldsymbol{I}_{(\boldsymbol{\omega},\boldsymbol{\psi})}, \Lambda_{(\boldsymbol{\omega},\boldsymbol{\psi})}\}\mathrm{d}x_{22}\mathrm{d}x_{12}\mathrm{d}x_{21}\mathrm{d}x_{11}.$$

As we have seen, we do not know the values of the final row and column of matrix $\Lambda_{(\boldsymbol{\omega},\boldsymbol{\psi})}$ or the final element of the vector $\boldsymbol{I}_{(\boldsymbol{\omega},\boldsymbol{\psi})}$. But, the fact mentioned in Theorem 2.2 becomes clear: this does not matter because the limits of integration corresponding to this variable are $(-\infty, \infty)$. Indeed, by the marginal distribution properties of the multivariate normal distribution, we need only consider one matrix $\Lambda_{(\boldsymbol{\omega},\boldsymbol{\psi})}$ and one set of vectors $\boldsymbol{I}_{l,(\boldsymbol{\omega},\boldsymbol{\psi})}$, exactly those given by the case $\boldsymbol{\omega} = (L, ..., L)^{\mathrm{T}}$. We denote these by Λ and $\boldsymbol{I}_l$ and set $\boldsymbol{I} = (\boldsymbol{I}_1^{\mathrm{T}}, ..., \boldsymbol{I}_L^{\mathrm{T}})$. Explicitly, we have

$$\mathrm{cov}(Z_{d_1 l_1}, Z_{d_2 l_2}) = \Lambda_{d_1+(D-1)(l_1-1), d_2+(D-1)(l_2-1)},$$
$$= \frac{1}{2}\left(\frac{l_1}{l_2}\right)^{1/2}(1 + \delta_{d_1 d_2}),$$
$$I_{dl} = \frac{ln}{2\sigma_e^2},$$

for any $d, d_1, d_2 \in \{1, ..., D-1\}$ and $l_1 \leq l_2, l, l_1, l_2 \in \{1, ..., L\}$.

Appendix B: Small sample size performance

For small sample sizes, simulation is required to accurately determine a design's performance. Because crossover trials are routinely conducted with small sample sizes, here we explore the impact this has upon the familywise error rate under the global null hypothesis.

We determined a design corresponding to the TOMADO example that would require only 12 patients in each of two stages, the smallest allowable maximum sample size for a group sequential crossover trial with $D = 4$ treatments initially, given our restrictions on n. Taking $\Delta = 0$ as an example, a trial with $n = 12$ and $L = 2$ with $f_1 = 0.768, f_2 = 2.036$, and $e_1 = 2.879$ to three decimal places would, using our multivariate normal calculations, have a maximal familywise error rate of $\alpha = 0.05$ under the global null hypothesis, and $\beta = 0.2$ for $\delta = 2.2$ when $\sigma_e^2 = 6.51$.

Ten thousand of these trials were simulated in order to ascertain the true probability of rejecting H_{0d} for some $d = 1, 2, 3$, when $\boldsymbol{\tau} = (0, 0, 0)^{\mathrm{T}}$. For simplicity, π_j was set to 0 for $j = 2, ..., D$, and μ_0 was set to 0. Incorporating non-zero period effects, however, would not be expected to greatly affect the results.

Whitehead et al. (2009) proposed a quantile substitution procedure for adapting the boundaries of a sequential trial to be more suitable to the case of unknown variance. We additionally considered employing this procedure. Given that there is no consensus on how to determine the degrees of freedom when analyzing using linear mixed models, we took the degrees of freedom at any analysis to be the classical decomposition of degrees of freedom in balanced, multilevel analysis of variance designs (Pinheiro and Bates, 2009). Moreover, we assessed the performance of the

sequential design when the linear mixed model was fitted through either maximum likelihood or restricted error maximum likelihood estimation. Therefore, in total, these simulations were performed for each of four possible analysis procedures: maximum likelihood or restricted error maximum likelihood estimation, with or without boundary adjustment through quantile substitution.

Thus, for each simulated study, patient response data for each stage l were randomly generated according to the distribution implied by their allocated treatment sequence (assigned according to the rules of the trial design), using the function rmvnorm (Genz et al., 2016) in R. The between-person variance was set to $\sigma_b^2 = 10.12$, the value ascertained in the final analysis of the TOMADO trial data. Following this, our linear mixed model was fitted on all accumulated data (with either maximum likelihood or restricted error maximum likelihood estimation according to the particular analysis procedure being considered) and $Z_{dl} = \hat{\tau}_{dl}\hat{I}_{dl}^{1/2}$ determined for $d = 1$, 2, 3, where $\hat{I}_{dl}^{1/2}$ is the observed Fisher information for $\hat{\tau}_{dl}$. Then, each Z_{dl} was compared to e_l and f_l and our stopping rules were applied (with e_l and f_l adjusted using quantile substitution if the analysis procedure under consideration so dictated). If for some $d = 1$, 2, 3, $f_l \leq Z_{dl} < e_l$, the trial proceeded to the following stage and the process was repeated. In each instance, simulations in which H_{0d} was rejected for some $d = 1$, 2, 3 were recorded in order to ascertain true rejection rates.

The performance of these procedures is displayed in Table B.1. We observe that when maximum likelihood estimation is utilized and the boundaries are not adjusted using the procedure of Whitehead et al. (2009), there is substantial inflation in the familywise error rate under the global null hypothesis to 0.077. However, when restricted error maximum likelihood estimation is used, there is only negligible inflation if adjustment of the boundaries is employed. A program to perform this analysis is available. Its use is detailed in Appendix D.

Appendix C: Technical proofs

Lemma C.1. *Element pq of Σ_r^{-1} is given by*

$$\Sigma_{rpq}^{-1} = \frac{1}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \left\{ (\sigma_e^2 + r\sigma_b^2)\delta_{pq} - \sigma_b^2 \right\}. \tag{C.1}$$

Proof. We demonstrate this by verifying $\Sigma_{rps}\Sigma_{rsq}^{-1} = \delta_{pq}$. From the chosen covariance for a patient's responses, we have that element pq of Σ_r is given by

$$\Sigma_{rpq} = \sigma_b^2 \mathbf{Z}_{rps}\mathbf{Z}_{rsq}^{\mathrm{T}} + \sigma_e^2\delta_{pq} = \sigma_b^2 + \sigma_e^2\delta_{pq}, \tag{C.2}$$

where $\mathbf{Z}_{rpq}, (p = 1; q = 1, ..., r)$, is the pqth element of $\mathbf{Z}_r$, the random effects design matrix for a single individual when there are r treatments remaining. Then

$$\begin{aligned}
\Sigma_{rps}\Sigma_{rsq}^{-1} &= \frac{1}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \sum_{s=1}^{r} \left\{ \sigma_b^2 + \sigma_e^2\delta_{ps} \right\} \left\{ (\sigma_e^2 + r\sigma_b^2)\delta_{sq} - \sigma_b^2 \right\}, \\
&= \frac{1}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \sum_{s=1}^{r} \left\{ \sigma_b^2(\sigma_e^2 + r\sigma_b^2)\delta_{sq} - \sigma_b^4 + \sigma_e^2(\sigma_e^2 + r\sigma_b^2)\delta_{pq} \quad - \sigma_e^2\sigma_b^2\delta_{ps} \right\}, \\
&= \frac{1}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \left\{ \sigma_b^2(\sigma_e^2 + r\sigma_b^2) - r\sigma_b^4 + \sigma_e^2(\sigma_e^2 + r\sigma_b^2)\delta_{pq} - \sigma_e^2\sigma_b^2 \right\}, \\
&= \delta_{pq},
\end{aligned}$$

using $\sum_s \delta_{sq} = \sum_s \delta_{ps} = 1$.

Lemma C.2. *Take the vector of fixed effects $\boldsymbol{\beta}$ to be*

$$\boldsymbol{\beta} = (\mu_0, \pi_2, ..., \pi_D, \tau_2, ..., \tau_{D-1})^{\mathrm{T}}.$$

Then we have the following result:

$$\sum_{i=1}^{|S_r|} \frac{n}{|S_r|} X_{s_{ri}}^{\mathrm{T}} \Sigma_r^{-1} X_{s_{ri}} = n \begin{pmatrix} A & B^{\mathrm{T}} & \mathbf{0}_{1,D-r} & B^{\mathrm{T}} & \mathbf{0}_{1,D-r} \\ B & C & \mathbf{0}_{r-1,D-r} & E & \mathbf{0}_{r-1,D-r} \\ \mathbf{0}_{D-r,1} & \mathbf{0}_{D-r,r-1} & \mathbf{0}_{D-r,D-r} & \mathbf{0}_{D-r,r-1} & \mathbf{0}_{D-r,D-r} \\ B & E & \mathbf{0}_{r-1,D-r} & C & \mathbf{0}_{r-1,D-r} \\ \mathbf{0}_{D-r,1} & \mathbf{0}_{D-r,r-1} & \mathbf{0}_{D-r,D-r} & \mathbf{0}_{D-r,r-1} & \mathbf{0}_{D-r,D-r} \end{pmatrix}, \tag{C.3}$$

where $\mathbf{0}_{m,n}$ is a matrix of zeroes of dimension $m \times n$, and

$$A = \frac{r\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)},$$

$$B_{pq} = \frac{r\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \qquad\qquad (p = 1, ..., r-1; q = 1),$$

$$C_{pq} = \frac{1}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)}\{(\sigma_e^2 + r\sigma_b^2)\delta_{pq} - \sigma_b^2\} \quad (p = 1, ..., r-1; q = 1, ..., r-1),$$

$$E_{pq} = \frac{1}{r}\frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \qquad\qquad (p = 1, ..., r-1; q = 1, ..., r-1).$$

Proof. Denote the columns of $X_{s_{ri}}$ by

$$X_{s_{ri}} = \begin{pmatrix} 1_r & \Pi_{2r} & ... & \Pi_{Dr} & T_{1s_{ri}} & ... & T_{D-1s_{ri}} \end{pmatrix}.$$

Thus, Π_{jr} is the column corresponding to the period effect π_j, $T_{ds_{ri}}$ to the treatment effect τ_d, and 1_r to the intercept μ_0. Using this representation and Lemma C.1, we have

$$X_{s_{ri}}^{\mathrm{T}} \Sigma_r^{-1} X_{s_{ri}} = \begin{pmatrix} 1_r^{\mathrm{T}} \\ \Pi_{2r}^{\mathrm{T}} \\ \vdots \\ T_{D-1s_{ri}}^{\mathrm{T}} \end{pmatrix} \Sigma_r^{-1} \begin{pmatrix} 1_r & \Pi_{2r} & ... & T_{D-1s_{ri}} \end{pmatrix}$$

$$= \begin{pmatrix} 1_r^{\mathrm{T}}\Sigma_r^{-1}1_r & 1_r^{\mathrm{T}}\Sigma_r^{-1}\Pi_{2r} & ... & 1_r^{\mathrm{T}}\Sigma_r^{-1}T_{D-1s_{ri}} \\ \Pi_{2r}^{\mathrm{T}}\Sigma_r^{-1}1_r & \Pi_{2r}^{\mathrm{T}}\Sigma_r^{-1}\Pi_{2r} & ... & \Pi_{2r}^{\mathrm{T}}\Sigma_r^{-1}T_{D-1s_{ri}} \\ \vdots & \vdots & \ddots & \vdots \\ T_{D-1s_{ri}}^{\mathrm{T}}\Sigma_r^{-1}1_r & T_{D-1s_{ri}}^{\mathrm{T}}\Sigma_r^{-1}\Pi_{2r} & ... & T_{D-1s_{ri}}^{\mathrm{T}}\Sigma_r^{-1}T_{D-1s_{ri}} \end{pmatrix}.$$

Here, $\dim(\Pi_{jr}) = \dim(1_r) = \dim(T_{js_ri}) = r \times 1$, for all i and j. Therefore, $1_r^{\mathrm{T}}\Sigma_r^{-1}1_r, \Pi_{jr}^{\mathrm{T}}\Sigma_r^{-1}1_r,$ $T_{js_{ri}}^{\mathrm{T}}\Sigma_r^{-1}1_r, 1_r^{\mathrm{T}}\Sigma_r^{-1}\Pi_{jr}, 1_r^{\mathrm{T}}\Sigma_r^{-1}T_{js_{ri}}, \Pi_{jr}^{\mathrm{T}}\Sigma_r^{-1}\Pi_{kr}, T_{js_{ri}}^{\mathrm{T}}\Sigma_r^{-1}T_{ks_{ri}},$ and $\Pi_{jr}^{\mathrm{T}}\Sigma_r^{-1}T_{ks_{ri}}$ are scalars for all i, j, and k.

By the definition of being in period j, the vth element of Π_{jr} is given by

$$\Pi_{jrv} = \delta_{jv}.$$

Because complete block design sequences are used, the vth element of $T_{js_{ri}}$ is given by

$$T_{js_{ri}v} = \delta_{1\mathbb{I}_{js_{ri}v}},$$

where

$$\mathbb{I}_{js_{ri}v} = \begin{cases} 1 & \text{An individual on sequence } s_{ri} \text{ receives treatment } j \text{ in period } v, \\ 0 & \text{otherwise.} \end{cases}$$

We denote this non-zero element, if it exists, by t_j.

Now from the symmetry present in Σ_r^{-1}, using Lemma C.1 we have

$$\sum_{j,k} \Sigma_{rjk}^{-1} = \frac{r\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)},$$

and

$$\sum_j \Sigma_{rjk}^{-1} = \sum_k \Sigma_{rjk}^{-1} = \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)},$$

for all j and k. Therefore, we can determine the form of each of the scalar elements in the matrix above as follows:

$$\mathbf{1}_r^T \Sigma_r^{-1} \mathbf{1}_r = \sum_{j,k} \Sigma_{rjk}^{-1} = \frac{r\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)},$$

$$\Pi_{jr}^T \Sigma_r^{-1} \mathbf{1}_r = \mathbf{1}_r^T \Sigma_r^{-1} \Pi_{jr} = \sum_k \Sigma_{rjk}^{-1} \mathbb{I}_{\{j\leq r\}} = \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \mathbb{I}_{\{j\leq r\}},$$

$$\boldsymbol{T}_{js_{ri}}^T \Sigma_r^{-1} \mathbf{1}_r = \mathbf{1}_r^T \Sigma_r^{-1} \boldsymbol{T}_{js_{ri}} = \sum_k \Sigma_{rt_jk}^{-1} \mathbb{I}_{\{j\leq r-1\}} = \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \mathbb{I}_{\{j\leq r-1\}},$$

$$\Pi_{jr}^T \Sigma_r^{-1} \Pi_{kr} = \Sigma_{rjk}^{-1} \mathbb{I}_{\{j\leq r\}} \mathbb{I}_{\{k\leq r\}} = \left\{ \frac{(\sigma_e^2 + r\sigma_b^2)\mathbb{I}_{\{j=k\}} - \sigma_b^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \right\} \mathbb{I}_{\{j\leq r\}} \mathbb{I}_{\{k\leq r\}},$$

$$\boldsymbol{T}_{js_{ri}}^T \Sigma_r^{-1} \boldsymbol{T}_{ks_{ri}} = \Sigma_{rt_jt_k}^{-1} \mathbb{I}_{\{j\leq r-1\}} \mathbb{I}_{\{k\leq r-1\}} = \left\{ \frac{(\sigma_e^2 + r\sigma_b^2)\mathbb{I}_{\{j=k\}} - \sigma_b^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \right\} \mathbb{I}_{\{j\leq r-1\}} \mathbb{I}_{\{k\leq r-1\}},$$

$$\Pi_{jr}^T \Sigma_r^{-1} \boldsymbol{T}_{ks_{ri}} = \Sigma_{rjt_k}^{-1} \mathbb{I}_{\{j\leq r\}} \mathbb{I}_{\{k\leq r-1\}} = \left\{ \frac{(\sigma_e^2 + r\sigma_b^2)\mathbb{I}_{\{j=t_k\}} - \sigma_b^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \right\} \mathbb{I}_{\{j\leq r\}} \mathbb{I}_{\{k\leq r-1\}}.$$

As a final step, we must compute the sum across sequences; that is, over the index i. We can see instantly that we have confirmed the elements proposed to be 0 in $\sum_{i=1}^{|S_r|} n\boldsymbol{X}_{s_{ri}}^T \Sigma_r^{-1} \boldsymbol{X}_{s_{ri}}/|S_r|$ are indeed so, and we therefore need only concentrate on the non-zero terms suggested: A, B, C, and E.

However, other than $\Pi_{jr}^T \Sigma_r^{-1} \boldsymbol{T}_{ks_{ri}}$, all of the elements above have been identified as independent of sequence s_{ri}. Therefore, computing the sum over the $s_{ri} \in S_r$ can be done easily and gives the forms proposed for A, B, and C in the statement of the lemma immediately, on multiplying through by $n/|S_r|$. But, by our imposed constraint that sequences be balanced for the period, we can also sum over the $\Pi_{jr}^T \Sigma_r^{-1} \boldsymbol{T}_{ks_{ri}}$

$$\sum_{i=1}^{|S_r|} \frac{n}{|S_r|} \Pi_{jr}^T \Sigma_r^{-1} \mathrm{T}_{ks_{ri}} = \frac{n}{|S_r|} \left(\frac{|S_r|}{r} \left[\frac{1}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \{\sigma_e^2 + (r-1)\sigma_b^2\} \right] + \left\{ 1 - \frac{|S_r|}{r} \right\} \left\{ \frac{-\sigma_b^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \right\} \right),$$

$$= \frac{n}{|S_r|} \left\{ \frac{|S_r|}{r} \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)} \right\},$$

$$= n\frac{1}{r} \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + r\sigma_b^2)},$$

since exactly $|S_r|/r$ patients receive each treatment at each time period. This confirms the form proposed for E, and the proof is complete.

Theorem C.1.   *(Theorem 2.1 from Section 2) Let $\boldsymbol{\beta} = (\mu_0, \pi_2, ..., \pi_D, \tau_1, ..., \tau_{D-1})^T$. Consider an analysis to be performed after some number of stages l. Then*

1. We have

$$\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD-1} = 0, L_{lD} = l) = \left(\frac{\ln}{|S_D|} \sum_{i=1}^{|S_D|} \boldsymbol{X}_{S_{Di}}^T \boldsymbol{\Sigma}_D^{-1} \boldsymbol{X}_{S_{Di}} \right)^{-1},$$

$$= \frac{1}{\ln} \begin{pmatrix} \boldsymbol{F} & \boldsymbol{G}^T & \boldsymbol{G}^T \\ \boldsymbol{G} & \boldsymbol{H} & \boldsymbol{0}_{D-1,D-1} \\ \boldsymbol{G} & \boldsymbol{0}_{D-1,D-1} & \boldsymbol{H} \end{pmatrix}, \qquad \text{(C.4)}$$

where

$$F = \sigma_b^2 + \frac{2D - 1}{D} \sigma_e^2,$$

$$\boldsymbol{G}_{pq} = -\sigma_e^2 \qquad\qquad (p = 1, ..., D - 1; q = 1),$$

$$\boldsymbol{H}_{pq} = \sigma_e^2 (1 + \delta_{pq}) \qquad (p = 1, ..., D - 1; q = 1, ..., D - 1).$$

1. If $q \geq 2$ is the largest integer such that $L_{lr} = 0$ for $r = 1, ..., q - 1$, then the covariance of the estimates of the fixed effects $\hat{\pi}_{2l}, ..., \hat{\pi}_{ql}, \hat{\tau}_{1l}, ..., \hat{\tau}_{q-1l}$ is identical to what it would be for $L_{l1} = ... = L_{lD-1} = 0$. Moreover, the covariance between the estimates of $\hat{\pi}_{2l}, ..., \hat{\pi}_{ql}, \hat{\tau}_{1l}, ..., \hat{\tau}_{q-1l}$ and the estimates of $\hat{\pi}_{q+1l}, ..., \hat{\pi}_{Dl}, \hat{\tau}_{ql}, ..., \hat{\tau}_{D-1l}$ is identical to what it would be for $L_{l1} = ... = L_{lD-1} = 0$.

Proof 1. We begin with the result for the case $L_{l1} = ... = L_{1D-1} = 0$. We demonstrate this by confirming

$$\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD-1} = 0, L_{lD} = l)^{-1} \text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD-1} = 0, L_{lD} = l) = \boldsymbol{I}_{2D-1}.$$

By Lemma C.2 we know that

$$\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD-1} = 0, L_{lD} = l)^{-1} = \ln \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{C} & \boldsymbol{E} \\ \boldsymbol{B} & \boldsymbol{E} & \boldsymbol{C} \end{pmatrix},$$

for

$$A = \frac{D\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)},$$

$$\boldsymbol{B} = \frac{D\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

$$\boldsymbol{C} = \frac{1}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} \sigma_e^2 + (D-1)\sigma_b^2 & -\sigma_b^2 & \cdots & -\sigma_b^2 \\ -\sigma_b^2 & \sigma_e^2 + (D-1)\sigma_b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\sigma_b^2 \\ -\sigma_b^2 & \cdots & -\sigma_b^2 & \sigma_e^2 + (D-1)\sigma_b^2 \end{pmatrix},$$

$$\boldsymbol{E} = \frac{1}{D} \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix},$$

with

$$\dim(\boldsymbol{B}) = (D-1) \times 1,$$
$$\dim(\boldsymbol{C}) = (D-1) \times (D-1),$$
$$\dim(\boldsymbol{E}) = (D-1) \times (D-1).$$

Thus, we must show

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^{\mathrm{T}} & \boldsymbol{B}^{\mathrm{T}} \\ \boldsymbol{B} & \boldsymbol{C} & \boldsymbol{E} \\ \boldsymbol{B} & \boldsymbol{E} & \boldsymbol{C} \end{pmatrix} \begin{pmatrix} \boldsymbol{F} & \boldsymbol{G}^{\mathrm{T}} & \boldsymbol{G}^{\mathrm{T}} \\ \boldsymbol{G} & \boldsymbol{H} & \boldsymbol{0}_{D-1,D-1} \\ \boldsymbol{G} & \boldsymbol{0}_{D-1,D-1} & \boldsymbol{H} \end{pmatrix} = \boldsymbol{I}_{2D-1}$$

or, on expanding,

$$\boldsymbol{AF} + 2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{G} = 1,$$
$$\boldsymbol{BG}^{\mathrm{T}} + \boldsymbol{CH} = \boldsymbol{I}_{D-1},$$
$$\boldsymbol{BG}^{\mathrm{T}} + \boldsymbol{EH} = \boldsymbol{0}_{D-1,D-1},$$
$$\boldsymbol{BF} + \boldsymbol{CG} + \boldsymbol{EG} = \boldsymbol{0}_{D-1,1},$$
$$\boldsymbol{AG}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}\boldsymbol{H} = \boldsymbol{0}_{1,D-1}.$$

Now

$$\boldsymbol{AF} + 2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{G} = \left(\frac{D\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \right)\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 \right) + 2\frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}^{\mathrm{T}} (-\sigma_e^2) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

$$= \left(\frac{D\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \right)\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 \right) - 2(D-1)\sigma_e^2 \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)},$$

$$= 1,$$

$$\boldsymbol{BG}^{\mathrm{T}} + \boldsymbol{CH} = \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (-\sigma_e^2) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}^{\mathrm{T}}$$

$$+ \frac{1}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} \sigma_e^2 + (D-1)\sigma_b^2 & -\sigma_b^2 & \dots & -\sigma_b^2 \\ -\sigma_b^2 & \sigma_e^2 + (D-1)\sigma_b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\sigma_b^2 \\ -\sigma_b^2 & \dots & -\sigma_b^2 & \sigma_e^2 + (D-1)\sigma_b^2 \end{pmatrix}$$

$$\times (\sigma_e^2) \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 2 \end{pmatrix},$$

$$= -\frac{\sigma_e^4}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 1 \end{pmatrix} + \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} 2\sigma_e^2 & \sigma_e^2 & \dots & \sigma_e^2 \\ \sigma_e^2 & 2\sigma_e^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_e^2 \\ \sigma_e^2 & \dots & \sigma_e^2 & 2\sigma_e^2 \end{pmatrix},$$

$$= \boldsymbol{I}_{D-1},$$

$$
\begin{aligned}
\boldsymbol{BG}^{\mathrm{T}} + \boldsymbol{EH} \quad &= \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}
(-\sigma_e^2)
\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}^{\mathrm{T}} \\
&\quad + \frac{1}{D}\frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 1 \end{pmatrix}
(\sigma_e^2)
\begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 2 \end{pmatrix}, \\
&= -\frac{\sigma_e^4}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 1 \end{pmatrix}
+ \frac{1}{D}\frac{\sigma_e^4}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
\begin{pmatrix} D & D & \dots & D \\ D & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & D \\ D & \ddots & D & D \end{pmatrix}, \\
&= \boldsymbol{0}_{D-1,D-1},
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{BF} + \boldsymbol{CG} + \boldsymbol{EG} \quad &= \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}
\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 \right) \\
&\quad + \frac{1}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
\begin{pmatrix} \sigma_e^2 + (D-1)\sigma_b^2 & -\sigma_b^2 & \dots & -\sigma_b^2 \\ -\sigma_b^2 & \sigma_e^2 + (D-1)\sigma_b^2 & \ddots & \vdots \\ \vdots & & \ddots & -\sigma_b^2 \\ -\sigma_b^2 & & \dots & -\sigma_b^2 \quad \sigma_e^2 + (D-1)\sigma_b^2 \end{pmatrix} \\
&\quad \times (\sigma_e^2)
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}
+ \frac{1}{D}\frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 1 \end{pmatrix}
(-\sigma_e^2)
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \\
&= \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 \right)
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}
- \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}
(\sigma_e^2 + \sigma_b^2)
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\
&\quad - \frac{1}{D}\frac{\sigma_e^4}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)}(D-1)
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \\
&= \boldsymbol{0}_{D-1,1},
\end{aligned}
$$

$$
\boldsymbol{A}\boldsymbol{G}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}\boldsymbol{H} = \left\{ \frac{D\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \right\} (-\sigma_e^2) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^{\mathrm{T}}
$$

$$
+ \left\{ \frac{\sigma_e^2}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \right\} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^{\mathrm{T}} (\sigma_e^2) \begin{pmatrix} 2 & 1 & \cdots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 2 \end{pmatrix},
$$

$$
= -\frac{D\sigma_e^4}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^{\mathrm{T}} + \frac{\sigma_e^4}{\sigma_e^2(\sigma_e^2 + D\sigma_b^2)} D \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^{\mathrm{T}},
$$

$$
= \boldsymbol{0}_{1,D-1},
$$

as required. Thus, the proposed matrix is indeed $\mathrm{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD-1} = 0, L_{lD} = l)$.

2. For the next part of the theorem, we re-order our vector of fixed effects such that

$$
\boldsymbol{\beta} = (\mu_0, \pi_2, \tau_1, \pi_3, \tau_2, ..., \pi_D, \tau_{D-1})^{\mathrm{T}}
$$

and thus the ordering of the columns of the $\boldsymbol{X}_{s_{ri}}$ is now

$$
\boldsymbol{X}_{s_{ri}} = \begin{pmatrix} 1_r & \Pi_{2r} & T_{1s_{ri}} & \Pi_{3r} & T_{2s_{ri}} & \cdots & \Pi_{Dr} & T_{D-1s_{ri}} \end{pmatrix}.
$$

We proceed by induction over the number of stages completed l for general D. Now, we assume that at the lth interim analysis, the statement of the theorem is true. Now, the covariance at this lth analysis is

$$
\mathrm{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1} = ... = L_{lq-1} = 0, L_{lq}, ..., L_{lD}) = \left(\sum_{r=q}^{D} L_r \frac{n}{|S_r|} \sum_{i=1}^{|S_r|} \boldsymbol{X}_{s_{ri}}^{\mathrm{T}} \Sigma_r^{-1} \boldsymbol{X}_{s_{ri}} \right)^{-1}
$$

$$
= \boldsymbol{A}^{-1}.
$$

It is this specifically that we assume follows the required condition. Additionally, we assume that this covariance matrix can be computed; that is, that $\boldsymbol{A}$ is invertible. We show that if we conduct another stage of the trial with t treatments remaining, $2 \le t \le q$, then the new covariance matrix

$$
\mathrm{cov}(\hat{\boldsymbol{\beta}}_{l+1}, \hat{\boldsymbol{\beta}}_{l+1} | L_{l+11}, ..., L_{l+1D}) = \left(\boldsymbol{A} + \frac{n}{|S_t|} \sum_{i=1}^{|S_t|} \boldsymbol{X}_{s_{ti}}^{\mathrm{T}} \Sigma_t^{-1} \boldsymbol{X}_{s_{ti}} \right)^{-1}
$$

$$
= (\boldsymbol{A} + \boldsymbol{B})^{-1}
$$

has the required property for $\hat{\pi}_{2l}, ..., \hat{\pi}_{tl}, \hat{\tau}_{1l}, ..., \hat{\tau}_{t-1l}$ as well as for $\hat{\pi}_{2l}, ..., \hat{\pi}_{tl}, \hat{\tau}_{1l}, ..., \hat{\tau}_{t-1l}$ and $\hat{\pi}_{t+1l}, ..., \hat{\pi}_{Dl}, \hat{\tau}_{tl}, ..., \hat{\tau}_{D-1l}$. Let

$$
\boldsymbol{T}_l = (\hat{\mu}_{0l}, \hat{\pi}_{2l}, \hat{\tau}_{1l}, ..., \hat{\pi}_{tl}, \hat{\tau}_{t-1l})^{\mathrm{T}},
$$

$$
\boldsymbol{T}_l' = (\hat{\pi}_{t+1l}, \hat{\tau}_{tl}, ..., \hat{\pi}_{Dl}, \hat{\tau}_{D-1l})^{\mathrm{T}},
$$

$$
\boldsymbol{W}_l = (\hat{\pi}_{2l}, \hat{\tau}_{1l}, ..., \hat{\pi}_{tl}, \hat{\tau}_{t-1l})^{\mathrm{T}}.
$$

Denote $\dim(\boldsymbol{T}_l) = |\boldsymbol{T}_l|$ and similarly for $\boldsymbol{T}_l'$ and $\boldsymbol{W}_l$.

By our assumptions, we can write

$$
\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{T}_l\boldsymbol{T}_l}^{-1} & \boldsymbol{A}_{\boldsymbol{T}_l\boldsymbol{T}_{l'}}^{-1} \\ \boldsymbol{A}_{\boldsymbol{T}_{l'}\boldsymbol{T}_l}^{-1} & \boldsymbol{A}_{\boldsymbol{T}_{l'}\boldsymbol{T}_{l'}}^{-1} \end{pmatrix},
$$

Where, for example, $\boldsymbol{A}_{\boldsymbol{T}_l\boldsymbol{T}_l}^{-1} = \mathrm{cov}(\boldsymbol{T}_l, \boldsymbol{T}_l | L_{l1}, ..., L_{lD})$ and with $\boldsymbol{A}_{\boldsymbol{T}_l\boldsymbol{T}_l}^{-1}$ and $\boldsymbol{A}_{\boldsymbol{T}_{l'}\boldsymbol{T}_l}^{-1} = (\boldsymbol{A}_{\boldsymbol{T}_l\boldsymbol{T}_{l'}}^{-1})^{\mathrm{T}}$ holding the required conditions for the fixed effects. Finally, as part of our inductive hypothesis we also assume that $\det(\boldsymbol{A}_{\boldsymbol{T}_l\boldsymbol{T}_l}^{-1}) > 0$.

With these definitions, our aim can then be stated to prove that

$$\text{cov}(\boldsymbol{W}_{l+1}, \boldsymbol{W}_{l+1} | L_{l+11}, ..., L_{l+1D}) = \frac{l}{l+1} A_{W_l W_l}^{-1};$$

that is, that the covariance between the fixed effects $\hat{\pi}_{2l}, ..., \hat{\pi}_{tl}$ and $\hat{\tau}_{1l}, ..., \hat{\tau}_{t-1l}$ is $l/(l+1)$ that of its form at interim analysis l and similarly that

$$\text{cov}(\boldsymbol{W}_{l+1}{}', \boldsymbol{W}_{l+1} | L_{l+11}, ..., L_{l+1D}) = \frac{l}{l+1} A_{W_{l'} W_l}^{-1}.$$

For brevity, from here we will write

$$\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | L_{l1}, ..., L_{lD}) = \text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l)$$

and similarly for $\boldsymbol{T}_l, \boldsymbol{T}_l{}'$, and $\boldsymbol{W}_l$ for any l.

We use the following identity, which requires only the invertibility of $\boldsymbol{A}$ to be valid (Henderson and Searle, 1981):

$$(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1}\{\boldsymbol{I} - \boldsymbol{U}\boldsymbol{C}\boldsymbol{V}\boldsymbol{A}^{-1}(\boldsymbol{I} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V}\boldsymbol{A}^{-1})^{-1}\}.$$

Note that we can write

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{T_{l+1}T_{l+1}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{|T_{l+1}|} \\ \boldsymbol{0} \end{pmatrix} \boldsymbol{B}_{T_{l+1}T_{l+1}} \begin{pmatrix} \boldsymbol{I}_{|T_{l+1}|} & \boldsymbol{0} \end{pmatrix} = \boldsymbol{U}\boldsymbol{C}\boldsymbol{V},$$

because our general form for $\sum_{i=1}^{|S_l|} \boldsymbol{X}_{S_{ti}}^{\mathrm{T}} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{X}_{S_{ti}}$ is only non-zero in a $|\boldsymbol{T}_l| \times |\boldsymbol{T}_l|$ block in the top left-hand corner by Lemma C.2. Therefore, provided that $\boldsymbol{A}$ is invertible, we can always invert $\boldsymbol{A} + \boldsymbol{B}$ to find the covariance matrix at the following interim analysis. Moreover, we have

$$\boldsymbol{B}\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{B}_{T_{l+1}T_{l+1}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}_{T_lT_l}^{-1} & \boldsymbol{A}_{T_lT_{l'}}^{-1} \\ \boldsymbol{A}_{T_{l'}T_l}^{-1} & \boldsymbol{A}_{T_{l'}T_{l'}}^{-1} \end{pmatrix},$$

$$= \begin{pmatrix} \boldsymbol{B}_{T_{l+1}T_{l+1}}\boldsymbol{A}_{T_lT_l}^{-1} & \boldsymbol{B}_{T_{l+1}T_{l+1}}\boldsymbol{A}_{T_lT_{l'}}^{-1} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix},$$

and

$$\boldsymbol{I}_{2D-1} + \boldsymbol{B}\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{B}_{T_{l+1}T_{l+1}}\boldsymbol{A}_{T_lT_l}^{-1} + \boldsymbol{I}_{|T_l|} & \boldsymbol{B}_{T_{l+1}T_{l+1}}\boldsymbol{A}_{T_lT_{l'}}^{-1} \\ \boldsymbol{0} & \boldsymbol{I}_{|T_{l'}|} \end{pmatrix}.$$

Now we need the formula (Henderson and Searle, 1981)

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix},$$

$$\Rightarrow \boldsymbol{M}^{-1} = \begin{pmatrix} (\boldsymbol{E} - \boldsymbol{F}\boldsymbol{H}^{-1}\boldsymbol{G})^{-1} & -\boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{H} - \boldsymbol{G}\boldsymbol{E}^{-1}\boldsymbol{F})^{-1} \\ -\boldsymbol{H}^{-1}\boldsymbol{G}(\boldsymbol{E} - \boldsymbol{F}\boldsymbol{H}^{-1}\boldsymbol{G})^{-1} & (\boldsymbol{H} - \boldsymbol{G}\boldsymbol{E}^{-1}\boldsymbol{F})^{-1} \end{pmatrix},$$

which implies

$$(\boldsymbol{I}_{2D-1} + \boldsymbol{B}\boldsymbol{A}^{-1})^{-1} = \begin{pmatrix} (\boldsymbol{B}_{T_{l+1}T_{l+1}}\boldsymbol{A}_{T_lT_l}^{-1} + \boldsymbol{I}_{|T_l|})^{-1} & -(\boldsymbol{B}_{T_{l+1}T_{l+1}}\boldsymbol{A}_{T_lT_l}^{-1} + \boldsymbol{I}_{|T_l|})^{-1}\boldsymbol{B}_{T_{l+1}T_{l+1}}\boldsymbol{A}_{T_lT_{l'}}^{-1} \\ \boldsymbol{0} & \boldsymbol{I}_{|T_{l'}|} \end{pmatrix}.$$

Note that to use this block matrix inversion formula we require $\boldsymbol{B}_{T_{l+1}T_{l+1}}\boldsymbol{A}_{T_lT_l}^{-1} + \boldsymbol{I}_{|T_l|}$ to be invertible. However, by the previous result of this theorem, only the variance of the intercept term in the form for $\boldsymbol{A}_{T_lT_l}^{-1}$ is dependent upon the value of t, so we have that

$$B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} = \frac{1}{l}\begin{pmatrix} V_1 & 0 \\ V_2 & I_{|T_l|-1} \end{pmatrix} = \frac{1}{l}\begin{pmatrix} V_1 & 0 \\ V_2 & I_{|W_l|} \end{pmatrix},$$

for some V_1, V_2; $\dim(V_1) = 1 \times 1, \dim(V_2) = |W_l| \times 1$. Now, by Lemma C.3, $\det(B_{T_{l+1}T_{l+1}}^{-1}) > 0$, which gives $\det(B_{T_{l+1}T_{l+1}}) = \{\det(B_{T_{l+1}T_{l+1}}^{-1})\}^{-1} > 0$. By assumption, $\det(A_{T_lT_l}^{-1}) > 0$, and thus

$$\det(B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1}) = \det(B_{T_{l+1}T_{l+1}})\det(A_{T_lT_l}^{-1}) = \left(\frac{1}{l}\right)^{|W_l|} V_1 > 0,$$

$$\Rightarrow V_1 > 0.$$

Therefore,

$$B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} + I_{|T_l|} = \frac{1}{l}\begin{pmatrix} V_1 + l & 0 \\ V_2 & (1+l)I_{|W_l|} \end{pmatrix},$$

$$\Rightarrow \det(B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} + I_{|T_l|}) > 0,$$

and $B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} + I_{|T_l|}$ is therefore invertible as required.

Now

$$\text{cov}(\hat{\boldsymbol{\beta}}_{l+1}, \hat{\boldsymbol{\beta}}_{l+1}) = \begin{pmatrix} A_{T_lT_l}^{-1} & A_{T_lT_{l'}}^{-1} \\ A_{T_{l'}T_l}^{-1} & A_{T_{l'}T_{l'}}^{-1} \end{pmatrix}$$

$$\times \left\{ I_{2D-1} - \begin{pmatrix} B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} & B_{T_{l+1}T_{l+1}}A_{T_lT_{l'}}^{-1} \\ 0 & 0 \end{pmatrix} \right.$$

$$\left. \begin{pmatrix} (B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} + I_{|T_l|})^{-1} & -(B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} + I_{|T_l|})^{-1}B_{T_{l+1}T_{l+1}}A_{T_lT_{l'}}^{-1} \\ 0 & I_{|T_{l'}|} \end{pmatrix} \right\}.$$

We thus have

$$\text{cov}(T_{l+1}, T_{l+1}) = A_{T_lT_l}^{-1}\{I_{|T_l|} - B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1}(B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} + I_{|T_l|})^{-1}\},$$

$$= A_{T_lT_l}^{-1}[I_{|T_l|} - \{I_{|T_l|} + (B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1})^{-1}\}^{-1}]$$

$$\text{cov}(T_{l+1}', T_{l+1}) = A_{T_{l'}T_l}^{-1}\{I_{|T_l|} - B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1}(B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} + I_{|T_l|})^{-1}\},$$

$$= A_{T_{l'}T_l}^{-1}[I_{|T_l|} - \{I_{|T_l|} + (B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1})^{-1}\}^{-1}],$$

by the identity $DC^{-1}(DC^{-1} + I)^{-1} = (I + (DC^{-1})^{-1})^{-1}$ (Henderson and Searle, 1981), which we can use as $B_{T_{l+1}T_{l+1}}A_{T_lT_l}^{-1} + I_{|T_l|}$ is invertible from earlier. Then

$$\text{cov}(T_{l+1}, T_{l+1}) = A_{T_lT_l}^{-1}\left(I_{|T_l|} - \left[I_{|T_l|} - \left\{\frac{1}{l}\begin{pmatrix} V_1 & 0 \\ V_2 & I_{|W_l|} \end{pmatrix}\right\}^{-1}\right]^{-1}\right),$$

$$= A_{T_lT_l}^{-1}\left[I_{|T_l|} - \left\{I_{|T_l|} - l\begin{pmatrix} V_1 & \mathbf{0} \\ V_2 & I_{|W_l|} \end{pmatrix}^{-1}\right\}^{-1}\right],$$

$$= A_{T_lT_l}^{-1}\left[I_{|T_l|} - \left\{I_{|T_l|} - l\begin{pmatrix} V_1^{-1} & \mathbf{0} \\ -I_{|W_l|}V_2V_1^{-1} & I_{|W_l|}^{-1} \end{pmatrix}\right\}^{-1}\right],$$

$$= A_{T_l T_l}^{-1} \left\{ I_{|T_l|} - \begin{pmatrix} 1 + lV_1^{-1} & 0 \\ -I_{|W_l|}V_2V_1^{-1} & (1+l)I_{|W_l|}^{-1} \end{pmatrix}^{-1} \right\},$$

$$= A_{T_l T_l}^{-1} \times \left\{ I_{|T_l|} - \begin{pmatrix} (1 + lV_1^{-1})^{-1} & 0 \\ -(1+l)^{-1}I_{|W_l|} - I_{|W_l|}V_2V_1^{-1}(1+lV_1^{-1})^{-1} & (1+l)^{-1}I_{|W_l|} \end{pmatrix} \right\},$$

$$= A_{T_l T_l}^{-1} \left\{ I_{|T_l|} - \begin{pmatrix} (1 + lV_1^{-1})^{-1} & 0 \\ (1+l)^{-1}I_{|W_l|}V_2V_1^{-1}(1+lV_1^{-1})^{-1} & (1+l)^{-1}I_{|W_l|} \end{pmatrix} \right\},$$

$$= A_{T_l T_l}^{-1} \begin{pmatrix} 1 - (1 + lV_1^{-1})^{-1} & 0 \\ (1+l)^{-1}I_{|W_l|}V_2V_1^{-1}(1+lV_1^{-1})^{-1} & I_{|W_l|}(1+l)^{-1}I_{|W_l|} \end{pmatrix}.$$

Now

$$\det\{\mathrm{cov}(T_{l+1}, T_{l+1}|L_{l+11}, ..., L_{l+1D})\} = \det(A_{T_l T_l}^{-1})\det\{1 - (1 + lV_1^{-1})^{-1}\} \times \det\{I_{|W_l| - (1+l)^{-1}I_{|W_l|}}\} > 0,$$

thus $\det(A_{T_{l+1} T_{l+1}}^{-1}) > 0$ as required, and

$$\mathrm{cov}(W_{l+1}, W_{l+1}) = A_{W_l W_l}^{-1}\{I_{|W_l|} - (1+l)^{-1}I_{|W_l|}\},$$
$$= A_{W_l W_l}^{-1}\left\{\frac{l}{l+1}I_{|W_l|}\right\},$$
$$= \frac{l}{l+1}A_{W_l W_l}^{-1}.$$

Similarly,

$$\mathrm{cov}(W_{l+1}', W_{l+1}) = A_{W_{l'} W_l}^{-1}\{I_{|W_l|} - (1+l)^{-1}I_{|W_l|}\},$$
$$= A_{W_{l'} W_l}^{-1}\left\{\frac{l}{l+1}I_{|W_l|}\right\},$$
$$= \frac{l}{l+1}A_{W_{l'} W_l}^{-1}.$$

Thus, the covariance of the fixed effects at the $(l+1)$th analysis has the desired property.

Now, as the base case, consider having completed one stage of the trial and proceeding to complete another with any number of treatments $t = 2, ..., D$ remaining. Then, in this instance,

$$A = \frac{n}{|S_D|}\sum_{i=1}^{|S_D|}X_{S_{Di}}^{\mathrm{T}}\Sigma_D^{-1}X_{S_{Di}}.$$

By the previous result of this theorem, this is indeed invertible and has the desired property; moreover, by Lemma C.3, $\det(A_{T_l T_l}^{-1}) > 0$. The proof is then complete.

Lemma C.3. *Consider the matrix from part (1) of Theorem C.1, the case*
$$L_{l1} = ... = L_{lD-1} = 0, L_{ld} = l$$

$$\left(\frac{ln}{|S_D|}\sum_{i=1}^{|S_D|}X_{S_{Di}}^{\mathrm{T}}\Sigma_D^{-1}X_{S_{Di}}\right)^{-1}.$$

Now consider restricting to the columns and rows corresponding to

$$T_l = (\hat{\mu}_{0l}, \hat{\pi}_{2l}, \hat{\pi}_{3l}, ..., \hat{\pi}_{tl}, \hat{\tau}_{1l}, \hat{\tau}_{2l}, ..., \hat{\tau}_{t-1l})^{\mathrm{T}},$$

for some $t = 2, ..., D$. The determinant of the matrix $\mathrm{cov}(T_l, T_l | L_{l1} = ... = L_{lD-1} = 0, L_{lD} = l)$ is strictly positive for any t.

Proof.   We have, by part (1) of Theorem C.1

$$\mathrm{cov}(T_l, T_l | L_{l1} = ... = L_{lD-1} = 0, L_{lD} = l) = \frac{1}{ln}\left(\begin{array}{cc} \sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 & M^{\mathrm{T}} \\ M & N \end{array}\right),$$

for

$$M = \left(\begin{array}{c} -\sigma_e^2 \\ \vdots \\ -\sigma_e^2 \end{array}\right),$$

$$N = \sigma_e^2\left(\begin{array}{cc} P & 0 \\ 0 & P \end{array}\right),$$

$$P = \left(\begin{array}{cccc} 2 & 1 & ... & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & ... & 1 & 2 \end{array}\right),$$

$$\dim(M) = (2t-2) \times 1,$$
$$\dim(N) = 2(t-1) \times 2(t-1),$$
$$\dim(P) = (t-1) \times (t-1).$$

Then

$$\det\left\{\frac{1}{ln}\left(\begin{array}{cc} \sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 & M^{\mathrm{T}} \\ M & N \end{array}\right)\right\} = \left(\frac{1}{ln}\right)^{2t-1}\det\left(\left(\begin{array}{cc} \sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 & M^{\mathrm{T}} \\ M & N \end{array}\right)\right),$$

$$= \left(\frac{1}{ln}\right)^{2t-1}\det(N)\det\left\{\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2\right) - M^{\mathrm{T}}N^{-1}M\right\}.$$

Now

$$N^{-1} = \frac{1}{\sigma_e^2}\left(\begin{array}{cc} P^{-1} & 0 \\ 0 & P^{-1} \end{array}\right),$$

$$\det(N) = \sigma_e^{4(t-1)}\det(P)^2.$$

We are therefore left to find P^{-1}. We have $PQ = I_{D-1}$ for

$$Q = \frac{1}{t}\left(\begin{array}{cccc} t-1 & -1 & ... & -1 \\ -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & ... & -1 & t-1 \end{array}\right).$$

Thus, $\boldsymbol{P}^{-1} = \boldsymbol{Q}$, and

$$\det\left\{\frac{1}{ln}\left(\begin{array}{cc} \sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 & \boldsymbol{M}^{\mathrm{T}} \\ \boldsymbol{M} & \boldsymbol{N} \end{array}\right)\right\} = \left(\tfrac{1}{ln}\right)^{2t-1}\det\left(\left(\begin{array}{cc} \sigma_b^2 + \frac{2D-1}{D}\sigma_e^2 & \boldsymbol{M}^{\mathrm{T}} \\ \boldsymbol{M} & \boldsymbol{N} \end{array}\right)\right),$$

$$= \left(\tfrac{1}{ln}\right)^{2t-1}\det(\boldsymbol{N})\det\left\{\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2\right) - \boldsymbol{M}^{\mathrm{T}}\boldsymbol{N}^{-1}\boldsymbol{M}\right\},$$

$$= \left(\tfrac{1}{ln}\right)^{2t-1}\sigma_e^{4(t-1)}\det(\boldsymbol{P})^2$$

$$\times\det\left\{\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2\right) - \sigma_e^2\left(\begin{array}{c}-1\\ \vdots \\ -1\end{array}\right)^{\mathrm{T}}\left(\begin{array}{cc}\boldsymbol{Q} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}\end{array}\right)\left(\begin{array}{c}-1\\ \vdots \\ -1\end{array}\right)\right\}.$$

This will be strictly positive provided that

$$\det\left\{\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2\right) - \sigma_e^2\left(\begin{array}{c}-1\\ \vdots \\ -1\end{array}\right)^{\mathrm{T}}\left(\begin{array}{cc}\boldsymbol{Q} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}\end{array}\right)\left(\begin{array}{c}-1\\ \vdots \\ -1\end{array}\right)\right\} > 0.$$

But

$$\det\left\{\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2\right) - \sigma_e^2\left(\begin{array}{c}-1\\ \vdots \\ -1\end{array}\right)^{\mathrm{T}}\left(\begin{array}{cc}\boldsymbol{Q} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}\end{array}\right)\left(\begin{array}{c}-1\\ \vdots \\ -1\end{array}\right)\right\}$$

$$= \det\left\{\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2\right) - \sigma_e^2\left(\begin{array}{c}-1/t\\ \vdots \\ -1/t\end{array}\right)^{\mathrm{T}}\left(\begin{array}{c}-1\\ \vdots \\ -1\end{array}\right)\right\},$$

$$= \det\left\{\left(\sigma_b^2 + \frac{2D-1}{D}\sigma_e^2\right) - \sigma_e^2\left(\frac{2t-2}{t}\right)\right\},$$

$$= \sigma_b^2 + \sigma_e^2\left(\frac{2}{t} - \frac{1}{D}\right) > 0,$$

because $t \leq D$. Thus, we have the required result.

Appendix D: Programs

D.1. Availability

The code described below is available to download from https://github.com/mjg211/article_code.

D.2. R

The R package groupSeqCrossover allows the determination and exploration of group sequential power family crossover trial designs. The function gsco is used to determine the design, taking inputs for the value of L, α, β, δ, σ_e^2, Δ, and sequence type ("latin" or "williams"). The function plot can then be used through S3 methods to plot power curves, the expected sample size, and the expected number of observations for varying true treatment effects. Moreover, the function simulategsco can be used to simulate group sequential crossover trials in order to assess their operating characteristics. This is especially useful in the case of small sample size designs. The code below, for example, identifies the discussed design for $\Delta = 0$ and then plots the expected sample size curve

    # Identify the design

```
Delta.0 <- gsco(Delta = 0)
# Plot E(N | tau_1 = … = tau_(D-1) = theta)
plot(Delta.0)
```

Similarly, the following code would allow the determination of the familywise error rate under the global null hypothesis (when analyzing using maximum likelihood estimation and using quantile substitution on the identified boundaries) of the design discussed in Appendix B:

```
simulate.fwer <- simulategsco(REML = F, adjust = T)
```

D.3. Matlab

In order to ease the understanding of the forms of equations (C.1) through (C.4), Matlab code employing symbolic algebra to return their forms is available. The user details a value for D, and four matrices are then returned. For example, consider the case $D = 4$

```
≫ [eqC_1, eqC_2, eqC_3, eqC_4] = groupSeqCrossoverMatrices(4);
```

eqC_2 contains the Σ_r for $r = 2, 3, 4$. Specifically

```
≫ eqC_2
eqC_2 = [b^2 + e^2, b^2, b^2, b^2]
[b^2, b^2 + e^2, b^2, b^2]
[b^2, b^2, b^2 + e^2, b^2]
[b^2, b^2, b^2, b^2 + e^2]
[b^2 + e^2, b^2, b^2, 0]
[b^2, b^2 + e^2, b^2, 0]
[b^2, b^2, b^2 + e^2, 0]
[0, 0, 0, 0]
[b^2 + e^2, b^2, 0, 0]
[b^2, b^2 + e^2, 0, 0]
[0, 0, 0, 0]
[0, 0, 0, 0]
```

From this we observe

$$\Sigma_2 = \begin{pmatrix} \sigma_e^2 + \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_e^2 + \sigma_b^2 \end{pmatrix},$$

$$\Sigma_3 = \begin{pmatrix} \sigma_e^2 + \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_e^2 + \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_e^2 + \sigma_b^2 \end{pmatrix},$$

$$\Sigma_4 = \begin{pmatrix} \sigma_e^2 + \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_e^2 + \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_e^2 + \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_e^2 + \sigma_b^2 \end{pmatrix}.$$

Note that we have to remove the rows and columns of zeroes from the Σ_r for $r < D$, and we use b and e for σ_b and σ_e respectively.

Similarly, eqC_1 contains the forms for Σ_r^{-1} for $r = 2, 3, 4$.

eqC_3 and eqC_4 correspond to the case $\boldsymbol{\beta} = (\mu_0, \pi_2, ..., \pi_D, \tau_1, ..., \tau_{D-1})^T$. eqC_3 contains

$$\sum_{i=1}^{|S_r|} \frac{n}{|S_r|} \boldsymbol{X}_{S_{ri}}^T \boldsymbol{\Sigma}_r^{-1} \boldsymbol{X}_{S_{ri}} \quad (r = 2, 3, 4).$$

Precisely, the first $2D - 1$ rows correspond to the case $r = 4$, the next $2D - 1$ to $r = 3$, and so forth.

Finally, eqC_4 contains the matrix $\text{cov}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_l | \boldsymbol{\omega} = (l, ..., l)^T, \boldsymbol{\psi})$.

Acknowledgments

Funding

References

Bender, R. and Lange, S. (2001). Adjusting for Multiple testing-when and how? *Journal of Clinical Epidemiology* 54: 343–349.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2011). *Applied Longitudinal Analysis*, NJ: Wiley.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2016). *mvtnorm: Multivariate Normal and t Distributions*, https://cran.r-project.org/web/packages/mvtnorm/.

Hauck, W. W., Preston, P. E., and Bois, F. Y. (1997). A Group Sequential Approach to Crossover Trials for Average Bioequivalence, *Journal of Biopharmaceutical Statistics* 7: 87–96.

Henderson, H. V. and Searle, S. R. (1981). On Deriving the Inverse of a Sum of Matrices, *SIAM Review* 23: 53–60.

Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*, Boca Raton: Chapman and Hall/CRC.

Jones, B. and Kenward, M. G. (2014). *Design and Analysis of Cross-Over Trials*, Boca Raton: Chapman and Hall/CRC.

Magirr, D., Jaki, T., and Whitehead, J. (2012). A Generalized Dunnett Test for Multi-Arm Multi-Stage Clinical Studies with Treatment Selection, *Biometrika* 99: 494–501.

The Mathworks Inc. (2016). *MATLAB 2016a*, Natick: Mathworks.

Mills, E. J., Chan, A. W., Wu, P., Vail, A., Guyatt, G. H., and Altman, D. G. (2009). Design, Analysis, and Presentation of Crossover Trials, *Trials* 10: 27.

Pampallona, S. and Tsiatis, A. A. (1994). Group Sequential Designs for One-sided and Two-Sided Hypothesis Testing with Provision for Early Stopping in Favor of the Null Hypothesis, *Journal of Statistical Planning and Inference* 42: 19–35.

Parmar, M. K. B., Carpenter, J., and Sydes, M. R. (2014). More Multiarm Randomised Trials of Superiority Are Needed, *Lancet* 384: 283–284.

Pinheiro, J. C. and Bates, D. (2009). *Mixed-Effects Models in S and S-PLUS*, New York: Springer.

Quinnell, T. G., Bennett, M., Jordan, J., Clutterbuck-James, A. L., Davies, M. G., Smith, I. E., Oscroft, N., Pittman, M. A., Cameron, M., Chadwick, R., Morrell, M. J., Glover, M. J., Fox-Rushby, J. A., and Sharples, L. D. (2014). A Crossover Randomised Controlled Trial of Oral Mandibular Advancement Devices for Obstructive Sleep Apnoea-Hypopnoea, *Thorax* 69: 938–945.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*, Vienna, Austria.

Senn, S. (2002). *Cross-Over Trials in Clinical Research*, Chichester: Wiley.

Wason, J. M. S. (2015). Multi-Arm Multi-Stage Designs for Clinical Trials with Treatment Selection, in *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*, O. Sverdlov, ed., pp. 389–410, Boca Raton: Chapman and Hall/CRC.

Wason, J. M. S. and Jaki, T. (2012). Optimal Design of Multi-Arm Multi-Stage Trials, *Statistics in Medicine* 31: 4269–4279.

Wason, J. M. S., Mander, A. P., and Thompson, S. G. (2012). Optimal Multistage Designs for Randomised Clinical Trials with Continuous Outcomes, *Statistics in Medicine* 31: 301–312.

Wason, J. M. S., Stecher, L., and Mander, A. P. (2014). Correcting for Multiple-Testing in Multi-Arm Trials: Is It Necessary and Is It Done?, *Trials* 15: 364.

Whitehead, J., Valdes-Marquez, E., and Lissmats, A. (2009). A Simple Two-Stage Design for Quantitative Responses with Application to a Study in Diabetic Neuropathic Pain, *Pharmaceutical Statistics* 8: 125–135.