Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

Adaptive Multivariate Global Testing

Giorgos MINAS, John A.D. ASTON, and Nigel STALLARD

We present a methodology for dealing with recent challenges in testing global hypotheses using multivariate observations. The proposed tests target situations, often arising in emerging applications of neuroimaging, where the sample size n is relatively small compared with the observations' dimension K. We employ adaptive designs allowing for sequential modifications of the test statistics adapting to accumulated data. The adaptations are optimal in the sense of maximizing the predictive power of the test at each interim analysis while still controlling the Type I error. Optimality is obtained by a general result applicable to typical adaptive design settings. Further, we prove that the potentially high-dimensional design space of the tests can be reduced to a low-dimensional projection space enabling us to perform simpler power analysis studies, including comparisons to alternative tests. We illustrate the substantial improvement in efficiency that the proposed tests can make over standard tests, especially in the case of n smaller or slightly larger than K. The methods are also studied empirically using both simulated data and data from an EEG study, where the use of prior knowledge substantially increases the power of the test. Supplementary materials for this article are available online.

KEY WORDS: Adaptive design; Multivariate test; Neuroimaging; Power analysis.

1. INTRODUCTION

In this work, we develop novel methodology for dealing with recent challenges in testing global hypotheses using multivariate observations. The classical approach for studying the problem, Hotelling's T^2-test (Hotelling 1931), can efficiently detect effects in every direction of the multivariate space when the sample size n is sufficiently large. However, in settings where n approaches or becomes smaller than the observation dimension K, T^2-test becomes respectively inefficient and inapplicable. This cost in efficiency, paid due to the need to search in every direction of the alternative space, seems particularly wasteful (but avoidable), if prior knowledge about the direction of the effect is available. Motivated by the latter settings, often arising in the increasingly important field of neuroimaging, we develop tests which are powerful in studies with $n \gg K$, but can also be efficient in situations where n is close to or smaller than K.

The proposed tests employ adaptive designs allowing for sequential modifications of the test statistic based on accumulated data. Such adaptive designs have straightforward but not exclusive application in clinical trials. A large literature on the subject (e.g., Bauer and Köhne 1994; Proschan and Hunsberger 1995; Lehmacher and Wassmer 1999; Müller and Schäfer 2001; Brannath, Posch, and Bauer 2002; Liu, Proschan, and Pledger 2002; Brannath, Gutjahr, and Bauer 2012) deals with the derivation of flexible procedures that allow for adaptations of the initial design without inflation of the Type I error rate. Some sequential designs (e.g., Denne and Jennison 2000) also permit design adaptations, but the latter need to be preplanned and independent of the interim test statistics. Adaptive designs are employed for many kinds of adaptations including sample size recalculation (Lehmacher and Wassmer 1999; Mehta and Pocock 2011), treatment or hypothesis selection (Kimani, Stallard, and Hutton 2009), and sample allocation to treatments (Zhu and Hu 2010). Despite the fact that many authors have stressed the potential for test statistic adaptation (e.g., Bauer and Köhne 1994; Bretz et al. 2009), there are only a few papers on the subject (Lang, Auterith, and Bauer 2000; Kieser, Schneider, and Friede 2002). Furthermore, various approaches for adaptive designs in multiple testing are available (see Bretz et al. 2009). These methods can efficiently detect few independently significant outcomes. However, it is well known that standard multiple testing methods (e.g., Bonferroni and Simes tests) become conservative and inefficient in settings, such as the typical neuroimaging studies, where strong dependencies and a large number of outcomes are present (D'Agostino and Russell 2005).

Similarly to the tests developed by O'Brien (1984), Läuter, Glimm, and Kropf (1998), and Minas et al. (2012), the proposed tests are based on linear combinations of the observation vectors. The crucial element in this approach is the weighting vector reducing the observation vectors to the scalar linear combinations. This defines the direction in which we decide to search for effects, and it can substantially affect both Type I and Type II error rate of the tests. O'Brien proposed deriving the weighting vectors under the assumption of uniform mean structure, while Läuter et al. showed that if the weighting vector is derived from the observation sums of products matrix, the Type I error is controlled and high power is attained under certain factorial structures. On the other hand, the tests in Minas et al. (2012) can attain high power levels independently of the mean and covariance structure but a part of the sample is used in a separate pilot study to learn the weighting vector.

In this work, linear combination test statistics, initially constructed using weighting vectors derived from prior information, are sequentially updated based on observed data at subsequent interim analyses in an adaptive design. Early termination of the study (due to early acceptance or rejection of the null hypothesis

Giorgos Minas (E-mail: *g.c.minas@warwick.ac.uk*) and John A. D. Aston (E-mail: *J.A.D.Aston@warwick.ac.uk*), Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK. Nigel Stallard, Division of Health Sciences, Warwick Medical School, University of Warwick, UK (E-mail: *n.stallard@warwick.ac.uk*). The authors would like to express their thanks to an AE and two referees for comments which helped improve the article. J.A.D.A. acknowledges partial support for this work from EPSRC Grant EP/K021672/1.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jasa*.

Published with License by Taylor & Francis
Journal of the American Statistical Association
June 2014, Vol. 109, No. 506, Theory and Methods
DOI: 10.1080/01621459.2013.870905

at an interim analyses) which is often of interest, especially in clinical trials, is also possible within our approach. Our methods provide a formal framework for optimally using prior information in constructing test statistics as has been suggested, but not implemented, in earlier papers (Pocock, Geller, and Tsiatis 1987; Läuter, Glimm, and Kropf 1996; Tang, Gnecco, and Geller 1989a).

While our tests maintain the two prime targets of adaptive designs, namely flexibility and Type I error control (Brannath et al. 2012), we also focus on attaining power optimality. Specifically, we employ the methods proposed by Spiegelhalter, Abrams, and Myles (2002) to derive optimal tests maximizing the predictive power of the test at each interim analysis. The methods of proofs can be useful in deriving optimal adaptive designs in more general settings. As we illustrate in Section 3, the results of Theorem 3.1 could be used to derive optimal designs for regression analysis for example.

The power performance of a multivariate test, lying in a possibly high-dimensional design space, can be hard to illustrate and interpret. Therefore, power analysis of multivariate tests is typically restricted to a limited part of the design space. We tackle this problem by reexpressing the $\mathcal{O}(K^2)$-dimensional design space as a lower dimensional easily interpretable space that is still sufficient to determine power. The crucial step here is to identify a measure quantifying the angular distance between the selected weighting vector and the optimal weighting vector and proving its sufficiency in computing power. These results provide wide understanding of the behavior of linear combination tests and allow us to extend earlier work on power analysis of single stage (Pocock, Geller, and Tsiatis 1987; Follmann 1996; Logan and Tamhane 2004) and sequential (Tang, Gnecco, and Geller 1989b; Tang, Geller, and Pocock 1993) linear combination tests, beyond low-dimensional observations or specific mean and covariance structures.

We perform extensive simulation studies to explore and compare the proposed and alternative single stage and sequential procedures throughout the design space. We show that linear combination tests outperform Hotelling's T^2-tests for the latter angular distance being below a certain value which, especially for sample sizes close to K, can be rather high. We further show that, in contrast to linear combination tests, such as O'Brien OLS test, with fixed weighting vectors, the adaptive linear combination tests can attain high power levels even in situations where the weighting vector selected at the planning stage is orthogonal to the true optimal (where, of course, a nonadaptive test would have zero power asymptotically). The advantages of the proposed tests are also illustrated through a real example taken from an EEG depression study (Läuter, Glimm, and Kropf 1996).

This article is organized as follows. In Section 2, we formulate the class of linear combination tests while in Section 3 we derive optimal, with respect to power, tests in this class. In Section 4, we present the results allowing us to characterize power based on low-dimensional summaries of the design parameters. In Section 5, we discuss the main results of extensive simulation studies performed using the latter results to explore power and compare the proposed tests with alternative global tests under various conditions, while in Section 6 we apply our procedures to an EEG depression study. Section 7 includes a short summary and discussion of the obtained results. Technical lemmas and proofs are provided in Supplementary Material A, while further illustrations of the simulation studies are provided in Supplementary Material B.

2. FORMULATION OF J-STAGE LINEAR COMBINATION TESTS

In the following, we formulate J-stage linear combination z and t-tests and define their error rate functions. We assume that the K-dimensional observation vectors $\boldsymbol{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ijK})^T$ of subjects $i = 1, 2, \ldots, n_j$, participating in stage j, $j = 1, 2, \ldots, J$, of the study, are independent and identically distributed Gaussian random variables

$$\boldsymbol{Y}_{ij} \sim N_K(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.1}$$

with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)^T$ and covariance matrix the positive definite $\boldsymbol{\Sigma} = (\sigma_{kk'})_{k,k'=1}^K$. In medical applications, the mean vector is often interpreted as the treatment effect. We wish to test the global null hypothesis of no treatment effect $H_0 : \boldsymbol{\mu} = \mathbf{0} = (0, 0, \ldots, 0)^T$ against the two-sided alternative $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$. Note that the methods which follow equally apply to the two-sample test with common covariance matrix, but we continue with the one-sample presentation to simplify notation.

The observation vectors $\boldsymbol{Y}_{ij}$, $i = 1, 2, \ldots, n_j$, of the jth stage are projected on the nonzero weighting vector $\boldsymbol{w}_j = (w_{j1}, w_{j2}, \ldots, w_{jK})^T$ and the projection magnitudes form the linear combinations $L_{ij} = \boldsymbol{w}_j^T \boldsymbol{Y}_{ij}$,$i = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, J$. The stagewise z and t statistics for testing H_0 against H_1 using the random sample of linear combinations L_{ij}, $i = 1, \ldots, n_j$, when $\boldsymbol{\Sigma}$ is either known or unknown, are respectively

$$Z_j = \frac{\bar{L}_j}{\sigma_j / n_j^{1/2}}, \qquad T_j = \frac{\bar{L}_j}{s_j / n_j^{1/2}}. \tag{2.2}$$

Here, σ_j^2 is the variance and $\bar{L}_j$, s_j^2 are the sample mean and sample variance of the linear combination L_j, respectively. Under assumption (2.1), the stagewise z and t statistics, Z_j, T_j, $j = 1, 2, \ldots, J$ are respectively normally and noncentrally t distributed, $Z_j \sim N(\bar{\theta}_j, 1)$ and $T_j \sim t_{\nu_j}(\bar{\theta}_j)$ with location parameter

$$\bar{\theta}_j = \theta_j \sqrt{n_j}, \quad \theta_j = \frac{\boldsymbol{w}_j^T \boldsymbol{\mu}}{\left(\boldsymbol{w}_j^T \boldsymbol{\Sigma} \boldsymbol{w}_j\right)^{1/2}}, \tag{2.3}$$

and $\nu_j = n_j - 1$. Under H_0, the z and t statistics are standard normal and Student's t random variables, that is, $Z_j \sim N(0, 1)$ and $T_j \sim t_{\nu_j}$. The two-sided stagewise p values of the z and t-tests are, respectively, $p_{z_j} = 2\Phi(-|Z_j|)$ and $p_{t_j} = 2\Psi_{\nu_j}(-|T_j|)$, where $\Phi(\cdot)$ and $\Psi(\cdot)$ are the cumulative distribution functions of the standard normal and Student's t-distribution with ν_j degrees of freedom, respectively.

At the jth analysis, $j = 1, 2, \ldots, J$, performed after the jth stage study, a combination function $C(\boldsymbol{p}_j)$ is used to combine the stagewise p values, $\boldsymbol{p}_j = (p_1, \ldots, p_j)$, of stages 1 to j (p_j either p_{z_j} or p_{t_j}). Rejection and acceptance critical values $\alpha_{1,j}$ and $\alpha_{0,j}$ ($0 \leq \alpha_{1,j} \leq \alpha < \alpha_{0,j} \leq 1$, $j = 1, 2, \ldots, J$) are used to decide whether to stop the study early and either reject or accept H_0, respectively. Specifically, the J-stage sequential design has

the following form:

$$\left.\begin{array}{ll} \text{At interim analysis} & \\ j = 1, 2, \ldots, J-1, & \\ \text{if } C(\boldsymbol{p}_j) \le \alpha_{1,j}\,, & \text{stop study and reject } H_0, \\ \text{if } C(\boldsymbol{p}_j) \ge \alpha_{0,j}\,, & \text{stop study and accept } H_0, \\ \text{otherwise,} & \text{continue to stage } j+1. \\ \text{At the final analysis J,} & \\ \text{if } C(\boldsymbol{p}_J) \le \alpha_{1,J}, & \text{stop study and reject } H_0, \\ \text{otherwise,} & \text{stop study and accept } H_0. \end{array}\right\} \quad (2.4)$$

Several combination functions are proposed in the literature. Bauer and Köhne (1994) suggested the use of Fisher's product combination function

$$C(\boldsymbol{p}_j) = \prod_{l=1}^{j} p_l, \quad (2.5)$$

while Lehmacher and Wassmer (1999) suggested the use of the inverse normal combination function. These two combination functions are the most commonly used in the literature (Bretz et al. 2009). The formulation and results which follow use the Fisher's product function in (2.5), but our results equally apply to other combination functions including the inverse normal.

Herein, we will refer to the J-stage tests with linear combination stagewise z and t-test statistics as the J-stage z and t-tests, respectively. The power function, that is, the probability to reject H_0, of the J-stage z or t-test is $\beta = \sum_{j=1}^{J} \beta_j$ where, $\beta_1 = \Pr(p_1 \le \alpha_{1,1})$, the first stage and

$$\beta_j = \Pr(C(\boldsymbol{p}_l) \in (\alpha_{1,l}, \alpha_{0,l})\ \forall\, l < j\ ;\ C(\boldsymbol{p}_j) \le \alpha_{1,j}), \quad (2.6)$$

the jth stage power functions, $j = 2, 3, \ldots, J$ (β, β_j either β_z, β_{z_j} or β_t, β_{t_j}, respectively). The boundaries $\alpha_{1,j}$, $\alpha_{0,j}$ are suitably chosen to satisfy the Type I error equation

$$\alpha = \alpha_{1,1} + \sum_{j=2}^{J} \int_{\alpha_{1,1}}^{\alpha_{0,1}} \int_{\alpha'_{1,2}}^{\alpha'_{0,2}} \cdots \int_{\alpha'_{1,j-1}}^{\alpha'_{0,j-1}} \alpha'_{1,j}\ \mathrm{d}p_{j-1} \ldots \mathrm{d}p_2 \mathrm{d}p_1, \quad (2.7)$$

where $\alpha'_{1,j} = \alpha_{1,j}/p_1 p_2 \ldots\ p_{j-1}$, $\alpha'_{0,j} = \alpha_{0,j}/p_1 p_2 \ldots\ p_{j-1}$ the conditional rejection and acceptance boundaries, respectively, of stage j, $j = 2, 3, \ldots, J$.

3. OPTIMAL J-STAGE z AND t-TESTS

The crucial element for these J-stage linear combination z and t-tests are the stage-wise weighting vectors $\boldsymbol{w}_j$. In this section we develop a methodology for optimally deriving these weighting vectors. The next lemma is the first step for computing the weighting vectors maximizing the power of the z and t-tests.

Lemma 3.1. Under (2.1), the power of the J-stage z and t-tests in (2.4) with combination function as in (2.5) is nondecreasing in the absolute value of θ_j in (2.3), $j = 1, 2, \ldots, J$.

Note that it can be straightforwardly shown that the above result hold for both one-sided stagewise tests and for the inverse normal combination function. The proof of the above lemma is surprisingly complex because for some range of values of θ_j an increase in $|\theta_j|$ decreases the probability to continue to the next stage and therefore the power of the subsequent stages, $\beta^{(j+1)} = \sum_{l=j+1}^{J} \beta_l$, decreases. In Supplementary Material A, we prove that even for these range of values of $|\theta_j|$, the decrease (in absolute value) in $\beta^{(j+1)}$ is bounded above by the increase in β_j.

The above result, except for being crucial for deriving Theorem 3.1, can also be useful for more general settings of adaptive designs. For example, Lemma 3.1 proves that if investigators wish to apply an adaptive z or t-test and are interested in maximizing the power of these procedures, they only need to sequentially maximize the location parameters of the stagewise test statistics separately. For instance, suppose that one is willing to conduct an adaptive design study to explore the relationship between an observation variable Y with a set of covariates X described by $\boldsymbol{Y}_j = \boldsymbol{X}_j \boldsymbol{b}_j + e_j$, $e_j \sim N_n(0, \sigma^2 \boldsymbol{I}_n)$, $j = 1, 2, \ldots, J$, independent. Then, our results prove that to maximize the power of the J-stage test with stagewise statistics the classical z and t statistics, with respect to the experimental design, it is sufficient to maximize $X_j^T X_j$, $j = 1, 2, \ldots, J$, which agrees with the standard practice of deriving optimal designs.

Considering the J-stage linear combination z and t-tests, Lemma 3.1 implies that to maximize the power of these tests with respect to the weighting vectors $\boldsymbol{w}_j$, it is sufficient to maximize the value of θ_j, $j = 1, 2, \ldots, J$. Using this result, we next derive the power-optimal weighting vector.

Theorem 3.1. Under (2.1), the power of the J-stage z and t-tests in (2.4) with combination function as in (2.5) are maximized with respect to the weighting vectors $\boldsymbol{w}_j$, $j = 1, 2, \ldots, J$, if and only if the latter are proportional to

$$\boldsymbol{\omega}^* = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}. \quad (3.1)$$

The last result provides the optimal, in terms of power, weighting vector for the J-stage linear combination tests $\boldsymbol{\omega}^*$. In Section 3.1, we show that $\boldsymbol{\omega}^*$, which expresses the multivariate treatment effect standardized with respect to the variance matrix $\boldsymbol{\Sigma}$, is central in characterizing the power of these tests. However, this optimal vector $\boldsymbol{\omega}^*$ depends on the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and therefore is also unknown. In the next section, we develop a methodology for selecting the weighting vectors $\boldsymbol{w}_j$ in practice. We propose using the information for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, available at each interim analysis, to optimally select $\boldsymbol{w}_j$, $j = 1, 2, \ldots, J$, where optimality is expressed here in terms of predictive power. The source of this information is the data collected from the stages completed before each interim analysis, but also prior information extracted from previous studies and expert clinical opinion. Predictive power allows the incorporation of this information into our procedures in a natural and plausible way. Note that, as we also explain in the next section, if Equation (2.7) is satisfied, the Type I error of these tests is controlled.

3.1 The Proposed z^* and t^* Tests

Prior information, $\mathcal{I}_0$, is used to inform standard conjugate multivariate priors for the observation mean and covariance matrix. We use the Gaussian–inverse-Wishart prior

$$\begin{aligned} (\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathcal{I}_0) &\sim N_K(\boldsymbol{m}_0, \boldsymbol{\Sigma}/n_0), \\ (\boldsymbol{\Sigma} \mid \mathcal{I}_0) &\sim \mathrm{IW}_{K\times K}\left(\nu_0, \boldsymbol{S}_0^{-1}\right), \end{aligned} \quad (3.2)$$

where $\boldsymbol{m}_0$ represents a prior estimate of the value of $\boldsymbol{\mu}$ and n_0 corresponds to the number of observations on which this prior estimate is based, while ν_0 and $\boldsymbol{S}_0$ respectively represent the degrees of freedom and the (positive definite) scale matrix of the inverse-Wishart prior.

Under this standard Bayesian model (see Gelman et al. 2004), the posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given the information set $\mathcal{I}_j = \{\mathcal{I}_0, \boldsymbol{y}_{(j)}\}$, consisting of the prior information $\mathcal{I}_0$ and the data collected up to the jth interim analysis $\boldsymbol{y}_{(j)} = [\boldsymbol{y}_1 \boldsymbol{y}_2 \dots \boldsymbol{y}_j]$ is $(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathcal{I}_j) \sim N_K(\boldsymbol{m}_j, \boldsymbol{\Sigma}/n_{(j)}), (\boldsymbol{\Sigma} \mid \mathcal{I}_j) \sim \mathrm{IW}_{K\times K}(\nu_j, \boldsymbol{S}_j^{-1})$. Here,

$$\begin{aligned} \boldsymbol{m}_j &= \frac{n_0\boldsymbol{m}_0 + n_{(j)}\bar{\boldsymbol{y}}_{(j)}}{n_0 + n_{(j)}}, \\ \boldsymbol{S}_j &= \boldsymbol{S}_0 + \nu_{(j)}\boldsymbol{S}_{\boldsymbol{y}_{(j)}} + \frac{n_0 n_{(j)}}{n_0 + n_{(j)}}\big(\bar{\boldsymbol{y}}_{(j)} - \boldsymbol{m}_0\big)\big(\bar{\boldsymbol{y}}_{(j)} - \boldsymbol{m}_0\big)^T, \end{aligned} \tag{3.3}$$

and $\nu_{(j)} = n_0 + n_{(j)} - 1$ with $n_{(j)} = n_1 + n_2 + \cdots + n_j$ and $\bar{\boldsymbol{y}}_{(j)} = \sum_{l=1}^{j}\sum_{i=1}^{n_j} \boldsymbol{y}_{il}/n_{(j)}$ respectively the sample size and sample mean of $\boldsymbol{y}_{(j)}$. Note that, due to the positive definiteness of the prior estimates $\boldsymbol{S}_0$, the posterior estimates $\boldsymbol{S}_j$ are also positive definite. Positive definiteness of $\boldsymbol{S}_0$ is required for our procedures to be applicable.

We wish to use this information to select the weighting vectors $\boldsymbol{w}_j$ optimally. Optimality here is expressed in terms of predictive power of the test. Predictive power (Spiegelhalter, Abrams, and Myles 2002) in the present context is derived by averaging the power of the J-stage z and t-tests over the distributions of the model parameters for a given information set. The predictive power for the first stage given the prior information set $\mathcal{I}_0$ is $B_1 = \Pr(p_1 < \alpha_{1,1} \mid \mathcal{I}_0)$ and for the jth stage, $j = 2, 3, \dots, J$, given the information set $\mathcal{I}_{j-1}$ is

$$B_j = \begin{cases} 1\,, & \mathcal{I}_{j-1} \text{ s.t. } C(\boldsymbol{p}_l) \le \alpha_{1,l} \\ & \quad \text{for } l \in \{1, 2, \dots, j-1\}, \\ 0\,, & \mathcal{I}_{j-1} \text{ s.t. } C(\boldsymbol{p}_l) \ge \alpha_{0,l} \\ & \quad \text{for } l \in \{1, 2, \dots, j-1\}, \\ \displaystyle\sum_{l=j}^{J} \Pr\left(C(\boldsymbol{p}_{l'}) \in (\alpha_{1,l'}, \alpha_{0,l'}), l' < l;\right. & \\ \qquad \left. C(\boldsymbol{p}_l) \le \alpha_{1,l} \mid \mathcal{I}_{j-1}\right), & \text{otherwise.} \end{cases} \tag{3.4}$$

The next result presents the weighting vectors that we suggest to use for the stagewise linear combination z and t-tests.

Theorem 3.2. Under (2.1) and (3.2), the jth stage predictive power, B_{z_j}, $j = 1, 2, \dots, J$, of the J-stage z-test in (3.4) is maximized with respect to the weighting vector $\boldsymbol{w}_j$ if and only if $\boldsymbol{w}_j$ is proportional to

$$\boldsymbol{w}_{z_j^*} = \boldsymbol{\Sigma}^{-1}\boldsymbol{m}_{j-1}. \tag{3.5}$$

Similarly, as we prove in Supplementary Material A, for $n_{(j-1)} \to \infty$, the jth stage predictive power, B_{t_j}, $j = 1, 2, \dots, J$, of the J-stage t-test in (3.4) is maximized with respect to the weighting vector $\boldsymbol{w}_j$ if and only if $\boldsymbol{w}_j$ is proportional to

$$\boldsymbol{w}_{t_j^*} = \boldsymbol{S}_{j-1}^{-1}\boldsymbol{m}_{j-1}, \tag{3.6}$$

where $\boldsymbol{m}_j$, $\boldsymbol{S}_j$ as in (3.3). The proposed J-stage tests, henceforth called (adaptive) z^* and t^*-tests, proceed as follows: for the jth analysis, $j = 1, 2, \dots, J$, (i) obtain $w_{z_j^*}$ or $w_{z_j^*}$ using (3.5) or (3.6), (ii) set w_j equal to $w_{z_j^*}$ or $w_{z_j^*}$ and compute the stage j statistic Z_j or T_j as in (2.2), (iii) calculate the stage j p-value, $p_{z_j} = 2\Phi(-|Z_j|)$ or $p_{t_j} = 2\Psi_{\nu_j}(-|T_j|)$, (iv) use all the observed p-values to perform the combination test in (2.4).

Importantly, the weighting vectors $\boldsymbol{w}_{z_j^*}$ and $\boldsymbol{w}_{t_j^*}$, given the prior information and the observed (if any) data $\boldsymbol{y}_{(j-1)}$, are fixed before collecting $\boldsymbol{y}_j$ and hence, under the standard conditions described in the following theorem, the Type I error of z^* and t^*-test, is preserved.

Theorem 3.3. Under (2.1) and for $\alpha_{1,j}, \alpha_{0,j}$, $j = 1, 2. \dots, J$ satisfying Equation (2.7), the Type I error of the z^* and t^*-tests is preserved at the nominal α level.

4. POWER CHARACTERIZATION (POC)

To study the performance of a test, we primarily need to explore the relationship between its power function and the design parameters. The latter might be, among others, the critical values, the sample size(s), and the model parameters. The critical values and the sample size(s) are scalar and therefore it is straightforward to visualize power even across all their possible values (e.g., using simulations). Their relation to power can then be easily described and understood. In univariate settings, this is also the case for the model parameters. However, in the multivariate setting, model parameters can be high-dimensional and therefore it is not practically feasible to visualize power over the whole design space. Power analysis is then typically restricted to a limited range of different structures of the model parameters. This might be sufficient for power analysis in specific settings, but it has obvious limitations in considering the general behavior of a testing procedure.

In the following, we encounter this problem in the context of linear combination tests and we provide a solution. We first consider the case of J-stage linear combination z and t-tests with fixed weighting vectors which, apart from providing a method for performing simple and efficient power analysis of tests such as the OLS test in O'Brien (1984, see Logan and Tamhane 2004; Pocock, Geller, and Tsiatis 1987; Tang, Geller, and Pocock 1993 for earlier work), also provides the intuition for the results considering the z^* and t^* tests. Note that in Section 4, the critical values and sample sizes (including the "prior" sample sizes) are assumed to be fixed and described by the design vector $\boldsymbol{d} = (\alpha_{0,1}, \alpha_{0,2}, \dots, \alpha_{0,J}, \alpha_{1,1}, \alpha_{1,2}, \dots, \alpha_{1,J}, \nu_0, n_0, n_1, \dots, n_J)$.

To provide greater insight to the subsequent results, it is also worth noting the joint distribution of the stagewise linear combination z statistics, Z_j, $j = 1, 2, \dots, J$, here for $J = 2$,

$$\begin{aligned} Pr\left(Z_1 \le z_1, Z_2 \le z_2\right) &= \int Pr(Z_1 \le z_1, Z_2 \le z_2 | \boldsymbol{y}_1)\, \mathrm{d}F(\boldsymbol{y}_1) \\ &= \int_{\{\boldsymbol{y}_1 : Z_1 \le z_1\}} \Phi(z_2 - \bar{\theta}_2(\boldsymbol{y}_1))\, \mathrm{d}F(\boldsymbol{y}_1), \end{aligned}$$

where $F(\boldsymbol{y}_1)$ the cdf of the first stage data, $\boldsymbol{y}_1$, and $\bar{\theta}_2(\boldsymbol{y}_1)$ the location parameter as in (2.3). The latter parameter is independent of $\boldsymbol{y}_1$, that is $\bar{\theta}_2(\boldsymbol{y}_1) = \bar{\theta}_2$, for the linear combination tests with fixed weighting vector, while for the adaptive z^* and t^* tests, $\bar{\theta}_2(\boldsymbol{y}_1)$ depends on $\boldsymbol{y}_1$ through the weighting vectors in (3.5) or (3.6), respectively. The next section focuses on

characterizing further the effect of the weighting vector, through the parameters $\bar{\theta}_j$, on the power function. Note that the power function can be easily derived from the joint distribution of the stagewise statistics by replacing z_j with suitable rejection or acceptance boundaries. In Supplementary Material A, we show that the above expression can be easily generalized to any $J > 1$ and that by replacing $\Phi(\cdot)$ with the cdf of the Student's t-distribution $\Psi(\cdot)$, we can easily derive the joint distribution of T_j, $j = 1, 2, \ldots, J$.

4.1 PoC for the J-Stage z and t-Tests With Fixed Weighting Vectors

To compute the power of the J-stage z and t-tests with fixed weighting vectors $\boldsymbol{w}_j = \boldsymbol{w}$, it is sufficient to know the design vector $\boldsymbol{d}$, as well as the stagewise location parameters θ_j in (2.3) which in this case are also fixed, that is, $\theta_j = \theta$. The latter can be reexpressed as

$$\theta = \frac{\boldsymbol{w}^T\boldsymbol{\mu}}{(\boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w})^{1/2}} = \frac{\tilde{\boldsymbol{w}}^T\tilde{\boldsymbol{\omega}}^*}{\|\tilde{\boldsymbol{w}}\|} = \|\,\tilde{\boldsymbol{\omega}}^*\,\|\cos(\text{ang}(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{\omega}}^*)), \quad (4.1)$$

where $\text{ang}(\tilde{\boldsymbol{w}}_j, \tilde{\boldsymbol{\omega}}^*)$ denotes the angle, in measured radians at the origin, between the vectors $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{\omega}}^*$. Here, $\tilde{\boldsymbol{w}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{w}$, $\tilde{\boldsymbol{\omega}}^* = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\omega}^* = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ are the standardized selected and optimal weighting vectors. In particular, the latter expresses the standardized multivariate treatment effect, generalizing the univariate ($K = 1$) standardized treatment effect μ/σ. Considering the weighting vector selection problem, the first equation in (4.1) implies that a weighting vector that increases the mean and/or decreases the variance of the linear combination gives higher power. The ambiguity in the latter expression becomes clearer by the standardization in the second equation which implies that the weighting vector selection can be expressed as a process of learning the standardized optimal weighting vector $\tilde{\boldsymbol{\omega}}^*$.

The last equation in (4.1) establishes two scalar measures which are sufficient to determine power. The first is the magnitude of $\tilde{\boldsymbol{\omega}}^*$, $\|\tilde{\boldsymbol{\omega}}^*\| = (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{1/2} = D_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$, which is the Mahalanobis distance between the distributions of the observation $\boldsymbol{Y}_{ij}$ under the null and the alternative hypotheses. The Mahalanobis distance is a generalization of the univariate signal-to-noise ratio and can be interpreted as a measure of deviation from the null hypothesis. In medical settings, it is a well-known global measure of the strength of the treatment effect. The second, $\cos(\text{ang}(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{\omega}}^*))$, is a measure of angular distance between the selected and the optimal weighting vector. It is a measure, in other words, of the distance of our weighting vector selection to the optimal choice. Under this representation, it becomes clear that, for fixed weighting vectors, the location parameter θ is equal to a measure ($D_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$) of the strength of the treatment effect scaled down by a measure ($\cos(\text{ang}(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{\omega}}^*))$) of the distance between the parameters and their prior estimates. The last results are formally stated in the next theorem.

Theorem 4.1. The design vector $\boldsymbol{d}$, the Mahalanobis distance $D_{\boldsymbol{\mu},\boldsymbol{\Sigma}} = (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{1/2}$ and the angle $\text{ang}(\tilde{\boldsymbol{\omega}}^*, \tilde{\boldsymbol{w}})$ between the vectors $\tilde{\boldsymbol{\omega}}^* = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ and $\tilde{\boldsymbol{w}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{w}$ are sufficient to determine the power function β of the J-stage linear combination z and t-tests with fixed weighting vectors $\boldsymbol{w}_j = \boldsymbol{w}$.

4.2 PoC for the z^*-Test

The sequential adaptation of the weighting vector increases the complexity within the relation between the power function and the design parameters. However, following similar methodology as above, analogous results can be derived. For this we use two steps, the first of which involves standardizing the procedure, similarly to (4.1), and the second establishing a rotation invariance property of the power function. The next lemma is a direct consequence of the standardization step summarizing $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{m}_0$ to the vectors $\tilde{\boldsymbol{\omega}}^*$ and $\tilde{\boldsymbol{w}}_{z_1^*}$.

Lemma 4.1. The design vector $\boldsymbol{d}$, the standardized optimal weighting vector $\tilde{\boldsymbol{\omega}}^* = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ and the standardized first-stage weighting vector $\tilde{\boldsymbol{w}}_{z_1^*}$ in (3.5) are sufficient to determine the power function β_{z^*}.

In the above result, we make use of the fact that the location parameter, $\theta_{z_j^*}$, of the z^*-test can be written as

$$\theta_{z_j^*} = \frac{\tilde{\boldsymbol{w}}_{z_j^*}^T\tilde{\boldsymbol{\omega}}^*}{\|\tilde{\boldsymbol{w}}_{z_j^*}\|}, \quad \tilde{\boldsymbol{w}}_{z_j^*} = \frac{n_0\tilde{\boldsymbol{w}}_{z_1^*} + n_{(j-1)}\tilde{\boldsymbol{w}}_{\bar{\boldsymbol{y}}_{(j-1)}}}{n_0 + n_{(j-1)}},$$
$$\tilde{\boldsymbol{w}}_{\bar{\boldsymbol{Y}}_{(j)}} = \boldsymbol{\Sigma}^{-1/2}\bar{\boldsymbol{Y}}_{(j)} \sim N_K\big(\tilde{\boldsymbol{\omega}}^*, \boldsymbol{I}/n_{(j)}\big) \quad (4.2)$$

which implies that the adaptive selection of the weighting vectors can be reexpressed as a procedure of adaptive estimation of the vector $\tilde{\boldsymbol{\omega}}^*$. Under this standardization, we can proceed to the rotation-invariance step which results in the next lemma.

Lemma 4.2. The power, β_{z^*}, of the z^*-test is invariant to rotations of the weighting vector $\tilde{\boldsymbol{w}}_{z_1^*}$ around the optimal weighting vector $\tilde{\boldsymbol{\omega}}^*$.

The idea behind Lemma 4.2 is that if $\tilde{\boldsymbol{w}}_{z_1^*}$ is rotated around $\tilde{\boldsymbol{\omega}}^*$, that is, $\tilde{\boldsymbol{w}}_{z_1^*}$ is replaced by $\dot{\boldsymbol{w}}_{z_1^*} = \boldsymbol{R}\tilde{\boldsymbol{w}}_{z_1^*}$, where $\boldsymbol{R}$ is a rotation matrix with rotation axis $\tilde{\boldsymbol{\omega}}_*$, the rejection region of the test is changed. However, the new rejection region is simply a rotation of the initial rejection region. That is, for each point say $\tilde{\boldsymbol{w}}_{\bar{\boldsymbol{y}}_{(j)}}$ in the initial rejection region, we can find a unique point, say $\dot{\boldsymbol{w}}_{\bar{\boldsymbol{y}}_{(j)}}$, in the rotated rejection region such that $\dot{\boldsymbol{w}}_{\bar{\boldsymbol{y}}_{(j)}} = \boldsymbol{R}\tilde{\boldsymbol{w}}_{\bar{\boldsymbol{y}}_{(j)}}$. Because the symmetrical Gaussian distribution of the observations $\tilde{\boldsymbol{w}}_{\bar{\boldsymbol{Y}}_{(j)}} \sim N_K(\tilde{\boldsymbol{\omega}}^*, \boldsymbol{I}/n_{(j)})$ remains unchanged under the rotation, the likelihood of the rejection region, that is, the power of the z^*-test, remains the same. The next theorem is direct consequence of Lemmas 4.1 and 4.2.

Theorem 4.2. The design vector $\boldsymbol{d}$, the Mahalanobis distance $D_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ and the angle $\text{ang}(\tilde{\boldsymbol{\omega}}^*, \tilde{\boldsymbol{w}}_{z_1^*})$ between the vectors $\tilde{\boldsymbol{\omega}}^*$ and $\tilde{\boldsymbol{w}}_{z_1^*}$ are sufficient to determine the power function β_{z^*}.

The above theorem states that the dependence of the power function on the model parameters and their prior estimates is described by simply a scalar measure of the strength of the treatment effect and a scalar measure of distance between the parameters and their prior estimates. It provides a sufficient description of power which is based on easily interpretable summaries and is considerably lower dimensional (importantly not depending on K, see Table 1). This allows us to perform power analysis of the adaptive J-stage z^*-test in a simple way potentially covering the whole design space.

4.3 PoC for the t^* Test

The need to estimate the unknown $\boldsymbol{\Sigma}$ increases substantially the dimension and the complexity of the design space. The sequential estimation of $\boldsymbol{\Sigma}$, in addition to $\boldsymbol{\mu}$, to obtain the weighting vectors $\boldsymbol{w}_{t_j^*}$, implies that the power analysis needs to account for both estimation procedures. For this, we write the weighting vector $\tilde{\boldsymbol{w}}_{t_j^*}$, $j = 1, 2, \ldots, J$ in (3.6) as

$$\begin{aligned}\tilde{\boldsymbol{w}}_{t_j^*} &= \boldsymbol{\Sigma}^{1/2}\boldsymbol{w}_{t_j^*} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{S}_{j-1}^{-1}\boldsymbol{m}_{j-1} = \boldsymbol{D}_j^{-1}\tilde{\boldsymbol{w}}_{z_j^*},\\ \boldsymbol{D}_j &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}_{j-1}\boldsymbol{\Sigma}^{-1/2}\end{aligned} \tag{4.3}$$

and $\tilde{\boldsymbol{w}}_{z_j^*}$ the jth standardized weighting vector of the z^*-test in (4.2). Here the $\boldsymbol{\Sigma}$-deviation matrix $\boldsymbol{D}_j$ is a measure of deviation of the estimate $\boldsymbol{S}_{j-1}$ in (3.3) from the parameter $\boldsymbol{\Sigma}$. The weighting vector $\tilde{\boldsymbol{w}}_{t_j^*}$ is then written as a product of the inverse of the matrix $\boldsymbol{D}_j$, that accounts for the estimation of $\boldsymbol{\Sigma}$, and the vector $\tilde{\boldsymbol{w}}_{z_j^*}$ which accounts for the estimation of $\boldsymbol{\mu}$, the latter taking $\boldsymbol{\Sigma}$ as known. We next follow the same steps as in Section 4.2 for deriving the PoC of the t^*-test. The standardization step results in the next lemma summarizing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and their prior estimates $\boldsymbol{m}_0$ and $\boldsymbol{S}_0$ to the vectors $\tilde{\boldsymbol{\omega}}^*$, $\tilde{\boldsymbol{w}}_{z_1^*}$ and the matrix $\boldsymbol{D}_1$ that have clear interpretation.

Lemma 4.3. The design vector $\boldsymbol{d}$, the matrix $\boldsymbol{D}_1$ in (4.3) and the vectors $\tilde{\boldsymbol{\omega}}^*$ and $\tilde{\boldsymbol{w}}_{z_1^*}$ are sufficient to determine the power function β_{t^*}.

Here, we use that the location parameter $\theta_{t_j^*}$ and the $\boldsymbol{\Sigma}$-deviation matrix $\boldsymbol{D}_j$ can be written as

$$\begin{aligned}\theta_{t_j^*} &= \frac{\tilde{\boldsymbol{w}}_{z_j^*}^T\boldsymbol{D}_j^{-1}\tilde{\boldsymbol{\omega}}^*}{\|\boldsymbol{D}_j^{-1}\tilde{\boldsymbol{w}}_{z_j^*}\|},\\ \boldsymbol{D}_j &= \boldsymbol{D}_1 + \nu_{(j-1)}\boldsymbol{S}_{\tilde{\boldsymbol{w}}_{\boldsymbol{y}_{(j-1)}}}\\ &\quad + \frac{n_0 n_{(j-1)}}{n_0 + n_{(j-1)}}\big(\tilde{\boldsymbol{w}}_{\bar{\boldsymbol{y}}_{(j-1)}} - \tilde{\boldsymbol{w}}_{z_1^*}\big)\big(\tilde{\boldsymbol{w}}_{\bar{\boldsymbol{y}}_{(j-1)}} - \tilde{\boldsymbol{w}}_{z_1^*}\big)^T,\end{aligned} \tag{4.4}$$

and that $\tilde{\boldsymbol{w}}_{z_j^*}$ can be written as the weighted average in (4.2). Here, $\boldsymbol{S}_{\tilde{\boldsymbol{w}}_{\boldsymbol{y}_{(j)}}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}_{\boldsymbol{y}_{(j)}}\boldsymbol{\Sigma}^{-1/2}$ is the covariance matrix of the sample $\tilde{\boldsymbol{w}}_{\boldsymbol{y}_{il}}$, $i = 1, 2, \ldots, n_l$, $l = 1, 2, \ldots, j$, where, importantly, $\tilde{\boldsymbol{w}}_{\boldsymbol{Y}_{il}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}_{il} \sim N_K(\tilde{\boldsymbol{\omega}}^*, \boldsymbol{I})$.

In a similar fashion to the previous section, we next establish the invariance of the power function under certain rotations of the prior estimates. For this, we define $\boldsymbol{V} = [\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_K]$ to be the matrix with columns the orthonormal eigenvectors of $\boldsymbol{D}_1$ and $\boldsymbol{\Lambda}_1 = \operatorname{diag}(\boldsymbol{\lambda}_1)$ the diagonal matrix with diagonal $\boldsymbol{\lambda}_1 = (\lambda_{11}, \lambda_{12}, \ldots, \lambda_{1K})^T$ the vector of the corresponding eigenvalues ($\lambda_{11} \geq \lambda_{21} \geq \cdots \geq \lambda_{1K} > 0$). We can then write $\boldsymbol{D}_1 = \boldsymbol{V}\boldsymbol{\Lambda}_1\boldsymbol{V}^T$, $\tilde{\boldsymbol{w}}_{z_1^*} = \boldsymbol{V}\boldsymbol{c}_{z_1^*}$, and $\tilde{\boldsymbol{\omega}}^* = \boldsymbol{V}\boldsymbol{c}^*$ where

$$\begin{aligned}c_{z_j^*,k} &= \cos(\operatorname{ang}(\boldsymbol{v}_k, \tilde{\boldsymbol{w}}_{z_j^*})), \quad c_k^* = \cos(\operatorname{ang}(\boldsymbol{v}_k, \tilde{\boldsymbol{\omega}}^*)),\\ &\quad k = 1, 2, \ldots, K.\end{aligned} \tag{4.5}$$

The rotation invariance property of the t^*-test is described in the next lemma.

Lemma 4.4. The power function β_{t^*} is invariant to simultaneous rotations of the vector $\tilde{\boldsymbol{w}}_{z_1^*}$ and the eigenvectors of the matrix $\boldsymbol{D}_1$ around the optimal weighting vector $\tilde{\boldsymbol{\omega}}^*$.

Table 1. Model and prior parameters of the z^* and t^*-tests, respectively, and their dimension

Parameters	Dimension	Parameters	Dimension
$\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{m}_0$	$(K^2 + 5K)/2$	$\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{m}_0, \boldsymbol{S}_0$	$K^2 + 3K$
$\tilde{\boldsymbol{\omega}}^*, \tilde{\boldsymbol{w}}_{z_1^*}$	$2K$	$\tilde{\boldsymbol{\omega}}^*, \tilde{\boldsymbol{w}}_{z_1^*}, \boldsymbol{D}_1$	$\frac{K^2+5K}{2}$
$D_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$, $\operatorname{ang}(\tilde{\boldsymbol{\omega}}^*, \tilde{\boldsymbol{w}}_{z_1^*})$	2	$\boldsymbol{c}^*, \boldsymbol{c}_{z_1^*}, \boldsymbol{\lambda}_1$	$3K$

The proof of Lemma (4.4) is similar to the proof of Lemma (4.2), albeit rather more complex. The next theorem is direct consequence of Lemmas 4.3 and 4.4.

Theorem 4.3. The design vector $\boldsymbol{d}$, the vector of eigenvalues $\boldsymbol{\lambda}_1$ of the matrix $\boldsymbol{D}_1$ in (4.3), and the vectors $\boldsymbol{c}_{z_1^*}$ and $\boldsymbol{c}^*$ in (4.5) are sufficient to determine the power function β_{t^*}.

As we can see in Table 1, the last result reduces the dimension of the design space of the t^*-test substantially, allowing us to explore power across the design space. While the design space, due to the covariance matrix estimation, still depends on K, it is reduced from order K^2 to order K.

Furthermore, this reduction provides an understanding of how the selection of the weighting vector affects power. This becomes clearer if we consider that $\theta_{t_j^*}$ in (4.4) can be written as

$$\theta_{t_j^*} = \frac{\boldsymbol{c}_{z_j^*}^T\boldsymbol{\Lambda}_j^{-1}\boldsymbol{c}^*}{\|\boldsymbol{\Lambda}_j^{-1}\boldsymbol{c}_{z_j^*}\|}, \quad j = 1, 2, \ldots, J,$$

where

$$\begin{aligned}\boldsymbol{c}_{z_j^*} &= \frac{n_0\boldsymbol{c}_{z_1^*} + n_{(j-1)}\boldsymbol{c}_{\bar{\boldsymbol{y}}_{(j-1)}}}{n_0 + n_{(j-1)}},\\ \boldsymbol{\Lambda}_j &= \boldsymbol{\Lambda}_1 + \nu_{(j-1)}\boldsymbol{S}_{\boldsymbol{c}_{\boldsymbol{y}_{(j-1)}}}\\ &\quad + \frac{n_0 n_{(j-1)}}{n_0 + n_{(j-1)}}\big(\boldsymbol{c}_{\bar{\boldsymbol{y}}_{(j-1)}} - \boldsymbol{c}_{z_1^*}\big)\big(\boldsymbol{c}_{\bar{\boldsymbol{y}}_{(j-1)}} - \boldsymbol{c}_{z_1^*}\big)^T.\end{aligned}$$

Here, $\boldsymbol{c}_{\bar{\boldsymbol{y}}_{(j)}}$ and $\boldsymbol{S}_{\boldsymbol{c}_{\boldsymbol{y}_{(j)}}}$ are the sample mean and sample covariance matrix of the transformed observation vectors $\boldsymbol{c}_{\boldsymbol{Y}_{(j)}} = [\boldsymbol{c}_{\boldsymbol{Y}_1}\boldsymbol{c}_{\boldsymbol{Y}_2}\ldots\boldsymbol{c}_{\boldsymbol{Y}_j}]$ with $\boldsymbol{c}_{\boldsymbol{Y}_l}$, $l = 1, 2, \ldots, j$, the matrix with columns $\boldsymbol{c}_{\boldsymbol{Y}_{il}} = \boldsymbol{V}_1^T\tilde{\boldsymbol{w}}_{\boldsymbol{Y}_{il}} \sim N_K(\boldsymbol{c}^*, \boldsymbol{I}), i = 1, 2, \ldots, n_j$. The last expressions show that the distance of the prior estimates $\boldsymbol{m}_0$, $\boldsymbol{S}_0$ to the model parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ can be expressed by the distances of the vectors $\boldsymbol{c}_{z_1^*}$ and $\boldsymbol{\lambda}_1^{-1} = (1/\lambda_{11}, \ldots, 1/\lambda_{1K})^T$ to $\boldsymbol{c}^*$, the latter directly reflected to power through $\theta_{t_j^*}$ (see the next section for more information).

In the special case of the first stage $\boldsymbol{\Sigma}$-deviation matrix being proportional to the identity matrix, that is, $\boldsymbol{D}_1 \propto \boldsymbol{I}$ ($\lambda_{11} = \lambda_{12} = \cdots = \lambda_{1K}$), as the next result shows, the design space can be reduced further.

Theorem 4.4. For $\boldsymbol{D}_1 = c^{-1}\boldsymbol{I}$, the design vector $\boldsymbol{d}$, the constant c, the Mahalanobis distance $D_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$, and the angle $\operatorname{ang}(\tilde{\boldsymbol{w}}_{z_1^*}, \tilde{\boldsymbol{\omega}}^*)$ are sufficient to determine the power function β_{t^*}.

The last theorem proves that, for $\boldsymbol{D}_1 \propto \boldsymbol{I}$, we can use the fact that the prior $\boldsymbol{\Sigma}$-deviation matrix $\boldsymbol{D}_1$ does not change the directions of $\tilde{\boldsymbol{w}}_{z_j^*}$'s, to show that the relation of β_{t^*} to the model parameters and their prior estimates can be described simply by the scalars $D_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ and $\operatorname{ang}(\tilde{\boldsymbol{w}}_{z_1^*}, \tilde{\boldsymbol{\omega}}^*)$. In the next section, we use

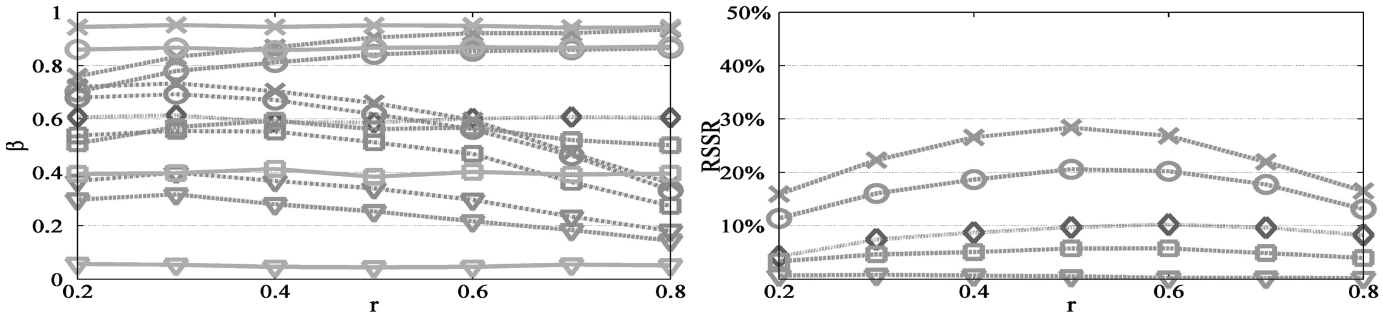

Figure 1. Power (left panel) and RSSR (right panel) versus sample allocation ratio. We plot the sequential χ^2-test (magenta $\cdot\diamond\cdot$) and the z^* (green $--$ line), sequential z (cyan $-$), and z^+ (orange $-\cdot$) tests with first stage/fixed/first step weighting vector 0 ($\times$), 30° ($\circ$), 60° ($\square$) and 90° ($\triangledown$) angle to the optimal. The remaining design parameters are $J = 2$, $K = 10$, $\alpha = 0.05$, $\alpha_{1,1} = 0.01$, $\alpha_{0,1} = 1$, $n_T = 60$, $n_0 = 0.5n_1$, $D_{\boldsymbol{\mu},\boldsymbol{\Sigma}} = 0.65$.

this result and the results of Theorems 4.2 and 4.3 to perform power analysis studies.

5. EMPIRICAL STUDIES

To explore properties of the adaptive z^* and t^*-tests as well as alternative global tests and to perform comparisons, we present empirical studies making use of the results in Theorems 4.2, 4.3, and 4.4.

In addition to z^* and t^*-tests, we consider linear combination z and t-tests with fixed weighting vectors, a class that includes the OLS z and t-test in O'Brien (1984). We also consider the likelihood-ratio χ^2 and Hotelling's T^2-test with statistics $\chi^2 = n\bar{\boldsymbol{Y}}\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{Y}}$ and $T^2 = n(n-K)\bar{\boldsymbol{Y}}\boldsymbol{S}_Y^{-1}\bar{\boldsymbol{Y}}/K(n-1)$ that follow the noncentral χ^2 and F distribution with K and $(K, n-K)$ degrees of freedom, respectively, and noncentrality parameter $D^2_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$. We consider both single stage and sequential J-stage designs for all these tests. Finally, the two-step, single-stage linear combination z^+ and t^+ tests proposed in Minas et al. (2012) are also considered. Note that the latter tests can be derived as special cases of the z^* and t^*-tests for $J = 2$, $(\alpha_{1,1}, \alpha_{0,1}) = (0, 1)$ and $C(\boldsymbol{p}_2) = p_2$.

A range of experiments are performed under different values of the design parameters. The power function of J-stage ($J > 1$) tests is not analytically tractable and therefore power is approximated by the rate of rejections in a large number of simulated replications, here $R = 10{,}000$, of a single experiment. Furthermore, to study the reduction in sample size due to early stopping of the study, we also empirically compute the rate of sample size reduction (RSSR),

$$\text{RSSR} = 100 \times \left(\frac{n_T - E(N)}{n_T}\right)\%,$$

where $n_T = n_1 + n_2 + \cdots + n_J$ the total sample size, N the sample size used for a single replication of the study and $E(N)$ its expected value. Note that single-stage tests have RSSR $= 0$, in contrast to sequential tests that allow for early stopping and thus have nonzero RSSR.

5.1 Simulation Data Examples

We next summarize the main results of a comprehensive study of the power behavior of the above tests in relation to the design parameters (more illustrations are included in Supplementary Material B). First, larger values of $D_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ and/or n_T result in higher power values for all tests considered, except the z and t-tests with fixed weighting vectors $\tilde{\boldsymbol{w}}$ orthogonal to $\tilde{\boldsymbol{\omega}}^*$ for which $\beta = \alpha$. Considering the prior sample size, the results indicate that for $n_0 \in (0.5n_1, 0.75n_1)$ the prior estimates become influential, but they do not dominate the accumulated data when selecting the weighting vector while larger values of n_0 enforces z^* and t^* to have more similar behavior to z and t-tests with fixed weighting vector. Furthermore, simulation examples confirm that larger values of the acceptance critical values $\alpha_{0,j}$ increase the power of multistage tests especially for larger potential power gain in subsequent stages, at the expense of less chance of early acceptance. Simulation examples also confirm that larger power is gained if larger rejection critical values $\alpha_{1,j}$ are allocated to stages with larger potential power gain, while the value of RSSR increases for larger $\alpha_{1,j}$ in early stages.

We also consider power behavior related to allocation of sample size to stages (Figure 1). For the sequential z and χ^2-test, the results show that higher power is achieved if sample allocation is analogous to α-rate allocation. The z^* and t^*-tests generally attain higher efficiency for close to balanced allocations. For $\tilde{\boldsymbol{w}}_{z_1^*}$ close to (far from) the optimal $\tilde{\boldsymbol{\omega}}^*$, slightly higher power is attained for assigning more sample to early (late) stages. Small to moderate allocation ratios r are more appropriate for the z^+ test since no α rate is spent in the first stage. Further, as in the χ^2-test, the z^* achieves higher RSSR for $r = 0.5$.

Before we proceed to comparisons, it is worth considering the impact of $\boldsymbol{\Sigma}$ being unknown and thus estimated on the performance of the t^*-test. First, in the case of $\boldsymbol{D}_1 \propto \boldsymbol{I}$ ($\boldsymbol{\lambda}_1 \propto \mathbf{1} = (1, 1, \ldots, 1)^T$), which as we show in Theorem 4.4 is somewhat easier case to consider, the $\boldsymbol{\Sigma}$ estimation variability is substantially reduced and thus we generally expect $\tilde{\boldsymbol{w}}_{t_j^*}$ to be closer to $\tilde{\boldsymbol{w}}_{z_j^*}$. On the other hand, if $\boldsymbol{D}_1 \not\propto \boldsymbol{I}$ ($\boldsymbol{\lambda}_1 \not\propto \mathbf{1}$), the direction of $\boldsymbol{\lambda}_1$ is more influential on $\tilde{\boldsymbol{w}}_{t_j^*}$ with the consequence being double-edged (see Figure 2). That is, compared to the situation of $\boldsymbol{\lambda}_1 \propto \mathbf{1}$, the distance of $\tilde{\boldsymbol{w}}_{t_j^*}$'s to optimal can be larger (left panel) but also smaller (right panel) depending on how close the direction of $\boldsymbol{\lambda}_1^{-1} = (1/\lambda_{11}, \ldots, 1/\lambda_{1K})^T$ is to the optimal direction $\boldsymbol{c}^*$.

Finally, it is useful to note that throughout our simulations of t^*-test, the $\cos(\text{ang}(\boldsymbol{c}^*, \boldsymbol{\Lambda}_1^{-1}\boldsymbol{c}_{z_1^*}))$ is shown to be a robust summary, albeit not sufficient (see Supplementary Material B, Figure 7, Section 2.1), of the distance between the model parameters

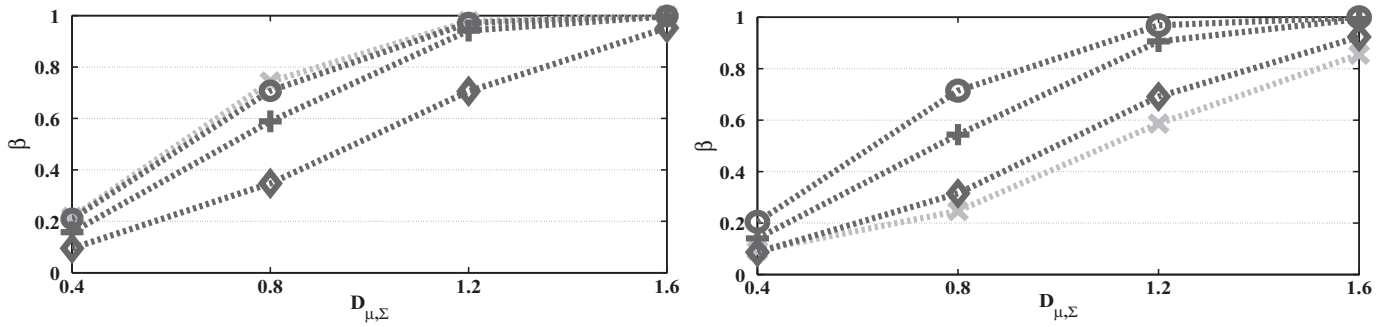

Figure 2. Power of the t^*-test versus Mahalanobis distance for various $\boldsymbol{c}^*, \boldsymbol{c}_{z_1^*}, \boldsymbol{\lambda}_1$. In the left panel, the vectors $\boldsymbol{c}^* = \boldsymbol{c}_{z_1^*} \propto \mathbf{1}$ while in the right panel $\boldsymbol{c}^* = \boldsymbol{e}_1 = (1, 0, \ldots, 0)^T$ and $\boldsymbol{c}_{z_1^*} \propto \mathbf{1}$ which, for $\boldsymbol{\lambda}_1 = \mathbf{1}$ (green $-\times-$ line), give $\varphi = \text{ang}(\boldsymbol{c}^*, \boldsymbol{\Lambda}_1^{-1}\boldsymbol{c}_{z_1^*}) = \text{ang}(\boldsymbol{c}^*, \boldsymbol{\lambda}_1^{-1}) = 0^\circ$ and 72°, respectively. In both panels, $\boldsymbol{\lambda}_1 \not\propto \mathbf{1}$ are also chosen to give $\varphi = 25^\circ$ (dark green $-\circ-$ line), 45° (dark green $-+-$ line) and 65° (dark green $-\diamond-$ line). The remaining design parameters are $J = 2$, $K = 10, \alpha = 0.05, \alpha_{1,1} = 0.01, \alpha_{0,1} = 1, n_T = 20, r = 0.5, n_0 = 0.75n_1, \nu_0 = n_0 - 1$.

and their prior estimates. For this reason, but also to reduce complexity, in the comparisons to follow, we focus on the case of $\boldsymbol{\lambda}_1 \propto \mathbf{1}$ (particularly, as we explain later on, in cases resembling the right panel of Figure 2), for various values of the summary $\cos(\text{ang}(\boldsymbol{c}^*, \boldsymbol{\Lambda}_1^{-1}\boldsymbol{c}_{z_1^*}))$.

In terms of comparisons, first note that, for fixed design parameters, single-stage tests attain higher power levels than multi-stage tests, nevertheless at the expense of not allowing for early stopping and thus not allowing for sample size reduction (RSSR = 0). Furthermore, it might be useful to emphasize that for fixed design parameters, the power of the linear combination test with weighting vector (either fixed or initial) set equal to the optimal weighting vector $\boldsymbol{\omega}^*$ attains the maximum power and provides an upper bound to all the other presented procedures, including Hotelling's T^2-test as proved in Minas et al. (2012) (Corollary 1). Compared to the z-tests with fixed weighting vectors $\boldsymbol{w}$, as we can see in Figure 3, the adaptive z^* lose some power for $\tilde{\boldsymbol{w}}$ $(= \tilde{\boldsymbol{w}}_{z_1^*})$ close to optimal but gains substantial amounts of power for $\tilde{\boldsymbol{w}}$ far from optimal, importantly avoiding the problem of z-tests having zero power for $\tilde{\boldsymbol{w}}$ orthogonal to optimal. This result emphasizes that, even though the power of the proposed tests remains sensitive to the prior information used to select the weighting vector, they are less sensitive to the initial selection of the weighting vector than the z and t-tests, where the weighting vector is fixed. The adaptive z^*-test also has substantially higher power to z^+ for small angles to the optimal and slightly lower power for large angles. Finally, the power of the single-stage and sequential χ^2-tests is approximately equal to the power of the z^*-test for $\tilde{\boldsymbol{w}}_{z_1^*}$ having respectively 60° and 45° angle with $\tilde{\boldsymbol{\omega}}^*$. Note that, as the results in Figure 3 confirm, all the considered tests control the Type I error at the nominal level $\alpha = 0.05$.

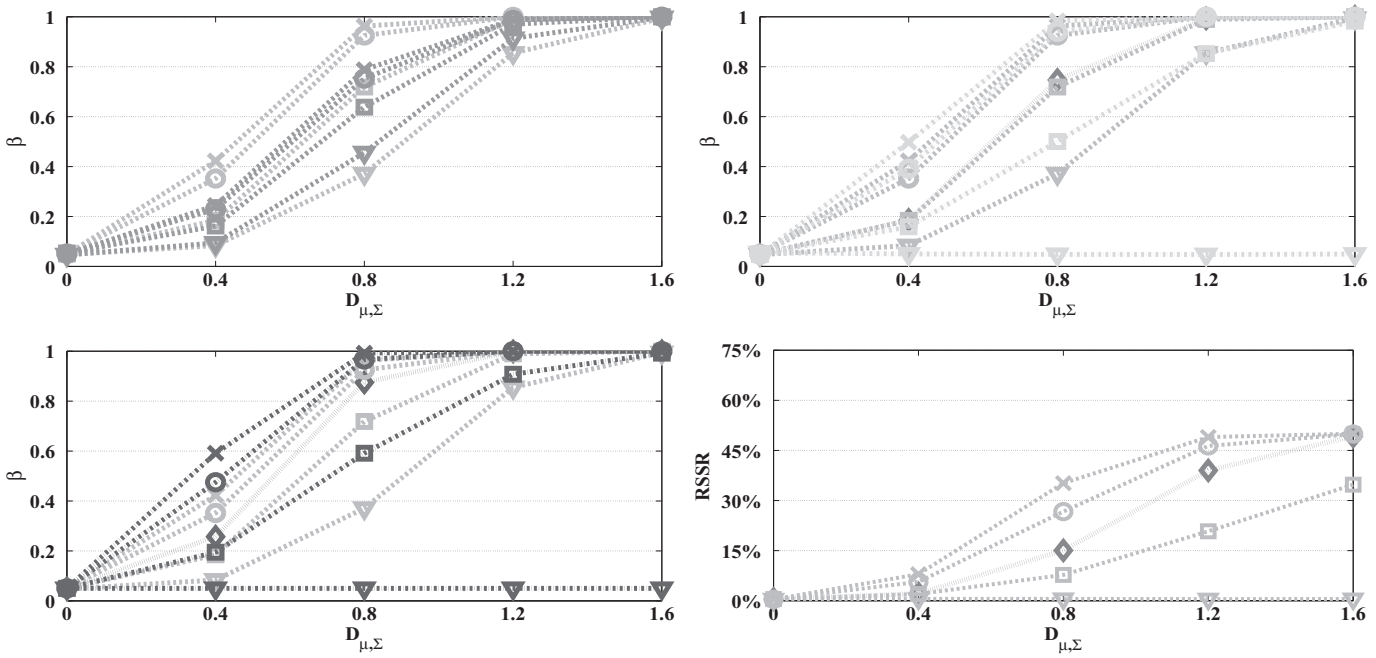

Figure 3. Power and RSSR versus Mahalanobis distance. We plot the z^*-test (green $--$) with the tests z^+ (orange $-.$) (up left), sequential z (cyan $-$) and χ^2 (magenta $\cdot\diamond\cdot$) (up right), single stage z (blue $-$) and χ^2 (red $\cdot\diamond\cdot$) (down left) and sequential χ^2 (down right). The linear combination $z^*/z/z^+$ tests are performed with first stage/fixed/first step weighting vectors having 0 ($\times$), 30° ($\circ$), 60° ($\square$), and 90° ($\triangledown$) angle to the optimal. The remaining design parameters are $J = 2$, $K = 10, \alpha = 0.05, \alpha_{1,1} = 0.01, \alpha_{0,1} = 1, n_T = 30, r = 0.5, n_0 = 0.75n_1, \nu_0 = n_0 - 1$.

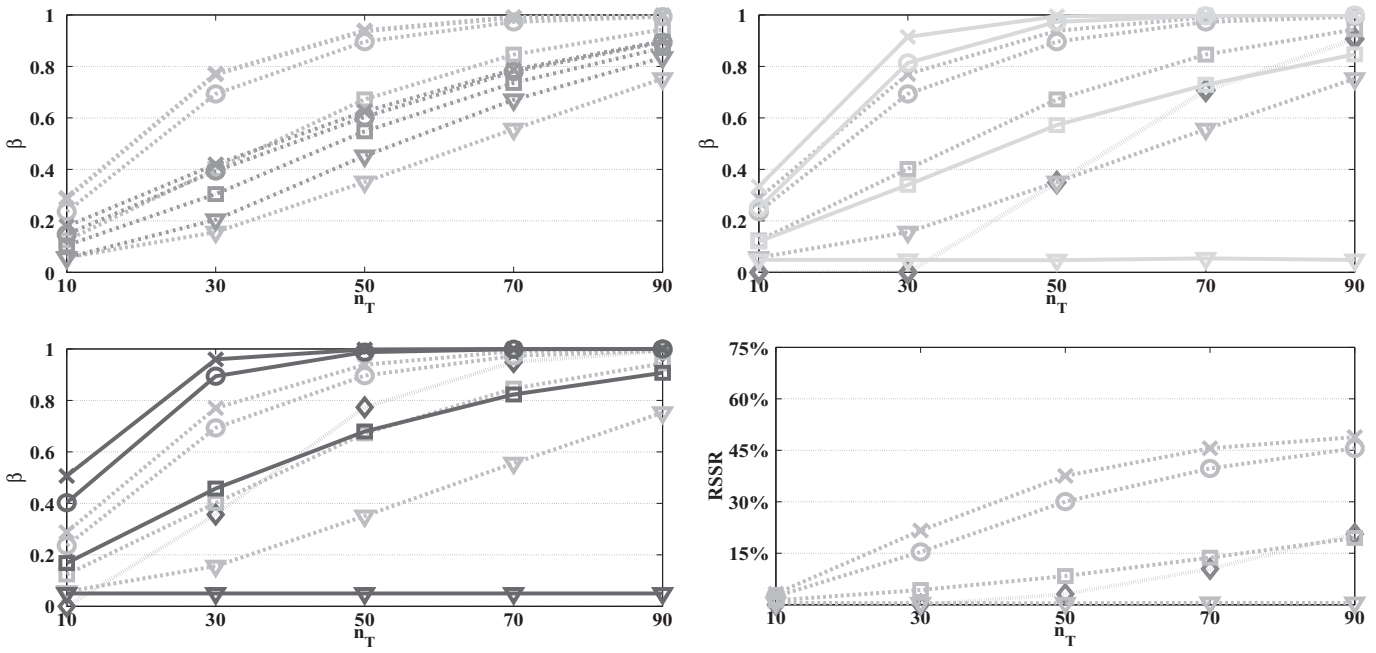

Figure 4. Power and RSSR versus the total sample size n_T. We plot the t^*-test (green $--$) with the tests, t^+ (orange $-.$) (up left), sequential t (cyan $-$) and T^2 (magenta $\cdot\diamond\cdot$) (up right), single stage t (blue $-$) and T^2 (red $\cdot\diamond\cdot$) (down left) and sequential T^2 (down right). The linear combination $t^*/t/t^+$ tests are performed with first stage/fixed/first step weighting vectors having 0 ($\times$), 30° ($\circ$), 60° ($\square$), and 90° ($\triangledown$) angle to the optimal. The remaining design parameters are $K = 15$, $J = 2$, $\alpha = 0.05$, $\alpha_{1,1} = 0.01$, $\alpha_{0,1} = 1$, $r = 0.5$, $n_0 = 6$, $\nu_0 = n_0 - 1$, $D_{\boldsymbol{\mu},\boldsymbol{\Sigma}} = 0.7$.

In the case of $\boldsymbol{\Sigma}$ unknown, we consider comparisons for the case of $\boldsymbol{D}_1 = \boldsymbol{I}$ which, using the results of Theorem 4.4, they can be performed in a similar way to the case of known $\boldsymbol{\Sigma}$. For the simulations in Figure 4, the case of $\boldsymbol{D}_1 = \boldsymbol{I}$ can be thought of as representative of $\boldsymbol{\lambda}_1^{-1}$ fairly distant to $\boldsymbol{c}^*$ (right panel of Figure 2), since we take $\boldsymbol{c}^* = \boldsymbol{e}_1$ resulting in $\cos(\text{ang}(\boldsymbol{c}^*, \boldsymbol{\lambda}_1^{-1})) = \sqrt{K}/K$ ($\cong 0.26$, angle 75°, for $K = 15$). As we would expect, the power of all tests is lower than their counterparts for $\boldsymbol{\Sigma}$ known (same design parameters), but the patterns of power difference across tests remain the same except from Hotelling's T^2 which in contrast to χ^2-test is highly dependent on the sample size.

As Figure 4 illustrates, for $n_T \leq K$ or n_T slightly larger than K (here, $n_T = 10-30$ for $K = 15$), T^2 is respectively inapplicable or very inefficient with power levels lower than the power of t^* even for angles close to orthogonal. As sample size becomes considerably bigger than K ($n_T > 50$), the power of T^2-test increases sharply to yield power levels analogous to the χ^2-test. For instance, for the design parameters in Figure 4, the single stage and sequential T^2-tests, likewise to the χ^2-test, have power close to the power of the t^* for angle 60° and 45°, respectively, for large sample sizes.

6. APPLICATION TO AN EEG STUDY

We consider applications to an electroencephalogram (EEG) study, the results of which are provided in Läuter, Glimm, and Kropf (1996). As Läuter et al. described, the data are collected from $n_T = 19$ depressive patients at the beginning and at the end of a six week therapy. For demonstration, $K = 9$ variables are used which represent the changes of the absolute theta power in channels 3–8, 17–19 of EEG during the therapy of each patient. In Table 2, we present the means, standard deviations, and correlation matrix of the data. Note that although an increase is indicated in all channels, none of them ($\min_k p_k = 0.04$) fall below the Bonferroni corrected threshold $\alpha/K \cong 0.0056$ at the $\alpha = 5\%$ significance level. Hotelling's T^2-test also fails to reject H_0 ($p_{T^2} = 0.261$). On the contrary, the SS and PC t-tests proposed by Läuter et al. reject H_0 at the 5% significance level ($p_{\text{SS}} = 0.0489$, $p_{\text{PC}} = 0.0487$).

We perform power analysis by setting the design parameters as in the above study, that is, $n_T = 19$, $K = 9$, $\boldsymbol{\mu} = \bar{y}$, $\boldsymbol{\Sigma} = \boldsymbol{S}_y$, $\alpha = 0.05$. For these design parameters, the power of Hotelling's T^2 is $\beta_{T^2} \cong 0.68$ ($D_{\boldsymbol{\mu},\boldsymbol{\Sigma}} = 1.15$). This is larger than the power of the SS and PC tests which are respectively $\beta_{t_{\text{SS}}} \cong 0.52$, $\beta_{t_{\text{PC}}} \cong 0.51$ (the contrasting results of the tests performed using these data are because of the different shape of the t and F distributions). The latter power values are very close to the power of the OLS t-test in O'Brien (1984), $\beta_{t_{\text{OLS}}} \cong 0.52$, which uses the uniform weighting vector $\boldsymbol{w}_{\text{OLS}} \propto \mathbf{1}$. This gives angle $\text{ang}(\tilde{\boldsymbol{w}}_{\text{OLS}}, \tilde{\boldsymbol{\omega}}^*) \cong 71°$. Taking into account that the single-stage t-test for a weighting vector equal to the optimal has power $\beta_t \cong 1$, we can easily see that there is considerable scope for improvement.

Since the study was performed, there has been considerable research into EEG studies on depressive patients. There is now literature (see, e.g., Davidson et al. 2002) indicating that left-frontal hypoactivation and right-frontal hyperactivation are present in such subjects. This would indicate that a nonuniform prior over these frontal regions should be used. Using prior information based on such evidence, the adaptive t^*-test can attain high power levels. For example, the prior estimates given in Table 2 are in agreement with the evidence in the literature and further, the prior correlation structure is set to be roughly coherent to the distances between the channels, that is, larger

Table 2. Means, standard deviations, correlations, and their prior estimates for the EEG depression study presented in Läuter, Glimm, and Kropf (1996)

ch.	3	4	5	6	7	8	17	18	19
$\bar{y}_k$	0.8710	1.5890	1.0370	1.1460	0.8510	0.8530	1.4220	0.7510	0.9950
$m_{0,k}$	0.5	3.50	1	2	2	2	2	2	2
s_{y_k}	2.9494	3.5121	2.3637	2.2490	2.2760	2.0706	3.2624	2.6382	2.3593
$s_{0,k}$	1.5	2.5	1	2	2	2	2	2	2
$R_0 \backslash R_y$	1	0.9262	0.8115	0.7959	0.5786	0.4902	0.9323	0.4896	0.5312
4	0.8	1	0.6270	0.7835	0.3357	0.4450	0.9313	0.2778	0.4892
5	0.8	0.7	1	0.7882	0.8492	0.7173	0.7347	0.7145	0.7611
6	0.7	0.8	0.7	1	0.6020	0.7924	0.8180	0.6334	0.7783
7	0.5	0.4	0.7	0.55	1	0.6155	0.4639	0.6833	0.5992
8	0.4	0.5	0.55	0.7	0.6	1	0.5177	0.5983	0.7833
17	0.9	0.9	0.75	0.75	0.45	0.45	1	0.4048	0.5711
18	0.45	0.45	0.65	0.65	0.7	0.7	0.5	1	0.4445
19	0.75	0.75	0.8	0.8	0.65	0.65	0.8	0.7	1

distances have smaller correlations, with larger correlations set at the highly active frontal regions (in accordance with the literature).

This prior estimate gives $\text{ang}(\tilde{\boldsymbol{w}}_{t_1^*}, \tilde{\boldsymbol{\omega}}^*) = 37.27^\circ$ which is much smaller than the angle under the uniform weighting vector. For a two-stage design ($J = 2$), with balanced sample allocation, $n_1 = 10$, $n_2 = 9$, and α allocation $\alpha_{1,1} = 0.01$, $a_2 = 0.0087$, no early acceptance allowed, $\alpha_{0,1} = 1$, prior sample size $n_0 = 7 = 0.7n_1$, $\nu_0 = 6$ (see previous section) and the remaining design parameters as the original study, the t^*-test has power $\beta_{t^*} \cong 0.84$ with RSSR $\cong 22.3\%$ ($E(N) \cong 15$). Substantial power improvement is also obtained over the t^+ which, for $n_0 = 6, n_1 = 13, n_2 = 6$ ($r = 0.3$) and the remaining design parameters as above, has power $\beta_{t^+} \cong 0.64$.

7. DISCUSSION

The methods developed in this work demonstrate that linear combination tests provide a substantial alternative to the classical Hotelling's T^2 global test, especially in the setting, commonly encountered in recent important applications of clinical neuroscience, of the available sample size n being small compared to the observation dimension K. It is also shown that adaptive linear combination tests provide power robustness across the set of alternative hypotheses since they can correct initial selections of the weighting vector which are far from the optimal selection. The adaptive J-stage z^* and t^*-tests achieve high power levels for large n, independently of the initial selection of weighting vector, but most importantly they can achieve high-power performance even if n is limited.

The proposed tests achieve optimality in the sense of maximizing the predictive power of the test at each interim analysis. Predictive power has been used for sample size calculation (O'Hagan and Stevens 2001), treatment selection (Kimani, Stallard, and Hutton 2009) and to select the component-wise significance levels in multiple testing (Westfall, Krishen, and Young 1998). It is a useful tool for incorporating prior information into the design of a study, particularly as such studies can often be viewed as a decision-making process. The application in Section 6 provides an example of a setting in which prior information is available and can substantially improve the performance of existing tests.

Optimality is attained in our methods without undermining the two main targets of adaptive designs: flexibility and test specificity. This allows for future developments of the proposed test to consider further optimal design adaptations. The use of other adaptive designs techniques, such as sample size reassessment, within our methodology can improve further the performance of the proposed tests.

The power characterization in Section 4 provides a tool for understanding and alleviating to some extent the complexities of multivariate tests especially those based on response dimension reductions. The possibly high-dimensional model parameters and their prior estimates are reduced to low-dimensional summaries which are still sufficient to compute power. Importantly, these summaries have interpretations directly related to the strength of the treatment effect and the effect of the dimension reduction on power. They provide a method for performing simple power analysis, but also understanding the behavior of linear combination tests.

The methods used to derive the power characterization are also interesting in their own right. They can be generally described by two steps: standardization and rotation invariance. The first standardization step is a prevalent technique for reexpressing statistical models in the standard deviation unit and eliminating correlations. Here, it allows us to reexpress the weighting vector selection, which involves estimating the unknown model parameters, as a procedure of learning a single vector, that is, the optimal weighting vector. The second step of establishing a rotation invariance property for the power function allows us to identify the measure quantifying the angular distance between the selected and the optimal weighting vector, reducing further the design space. The question whether these results can be derived under more relaxed modeling assumptions is an area of ongoing research.

SUPPLEMENTARY MATERIALS

Additional supplementary material is provided in the following documents:

Supplement A: Technical results Technical details, lemmas, and proofs.

Supplement B: Extended simulation examples Examples from the extensive simulation studies performed to study the power of the considered tests.

[Received April 2013. Revised October 2013.]

REFERENCES

Bauer, P., and Köhne, K. (1994), "Evaluation of Experiments With Adaptive Interim Analyses," *Biometrics*, 50, 1029–1041. [613,615]

Brannath, W., Gutjahr, G., and Bauer, P. (2012), "Probabilistic Foundation of Confirmatory Adaptive Designs," *Journal of the American Statistical Association*, 107, 824–832. [613,614]

Brannath, W., Posch, M., and Bauer, P. (2002), "Recursive Combination Tests," *Journal of the American Statistical Association*, 97, 236–244. [613]

Bretz, F., Koenig, F., Brannath, W., Glimm, E., and Posch, M. (2009), "Adaptive Designs for Confirmatory Clinical Trials," *Statistics in Medicine*, 28, 1181–1217. [613,615]

D'Agostino, R. B., and Russell, H. K. (2005), *Multiple Endpoints, Multivariate Global Tests*, New York: Wiley. [613]

Davidson, R. J., Pizzagalli, D., Nitschke, J. B., and Putnam, K. (2002), "Depression: Perspectives From Affective Neuroscience," *Annual Review of Psychology*, 53, 545–574. [621]

Denne, J. S., and Jennison, C. (2000), "A Group Sequential T-test With Updating of Sample Size," *Biometrika*, 87, 125–134. [613]

Follmann, D. (1996), "A Simple Multivariate Test for One-Sided Alternatives," *Journal of the American Statistical Association*, 91, 854–861. [614]

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Boca Raton, FL: Chapman & Hall. [616]

Hotelling, H. (1931), "The Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, 2, 360–378. [613]

Kieser, M., Schneider, B., and Friede, T. (2002), "A Bootstrap Procedure for Adaptive Selection of the Test Statistic in Flexible Two-Stage Designs," *Biometrical Journal*, 44, 641–652. [613]

Kimani, P. K., Stallard, N., and Hutton, J. L. (2009), "Dose Selection in Seamless Phase II/III Clinical Trials Based on Efficacy and Safety," *Statistics in Medicine*, 28, 917–936. [613,622]

Lang, T., Auterith, A., and Bauer, P. (2000), "Trendtests With Adaptive Scoring," *Biometrical Journal*, 42, 1007–1020. [613]

Läuter, J., Glimm, E., and Kropf, S. (1996), "New Multivariate Tests for Data With an Inherent Structure," *Biometrical Journal*, 38, 1–23. [614,621]

——— (1998), "Multivariate Tests Based on Left-Spherically Distributed Linear Scores," *The Annals of Statistics*, 26, 1972–1988. [613]

Lehmacher, W., and Wassmer, G. (1999), "Adaptive Sample Size Calculations in Group Sequential Trials," *Biometrics*, 55, 1286–1290. [613,615]

Liu, Q., Proschan, M. A., and Pledger, G. W. (2002), "A Unified Theory of Two-Stage Adaptive Designs," *Journal of the American Statistical Association*, 97, 1034–1041. [613]

Logan, B. R., and Tamhane, A. C. (2004), "On O'Brien's OLS and GLS Tests for Multiple Endpoints," *Lecture Notes-Monograph Series*, 47, 76–88. [614,616]

Mehta, C. R., and Pocock, S. J. (2011), "Adaptive Increase in Sample Size When Interim Results are Promising: A Practical Guide With Examples," *Statistics in Medicine*, 30, 3267–3284. [613]

Minas, G., Rigat, F., Nichols, T. E., Aston, J. A. D., and Stallard, N. (2012), "A Hybrid Procedure for Detecting Global Treatment Effects in Multivariate Clinical Trials: Theory and Applications to fMRI Studies," *Statistics in Medicine*, 31, 253–268. [613,619,620]

Müller, H.-H., and Schäfer, H. (2001), "Adaptive Group Sequential Designs for Clinical Trials: Combining the Advantages of Adaptive and of Classical Group Sequential Approaches," *Biometrics*, 57, 886–891. [613]

O'Brien, P. C. (1984), "Procedures for Comparing Samples With Multiple Endpoints," *Biometrics*, 40, 1079–1087. [613,616,619,621]

O'Hagan, A., and Stevens, J. W. (2001), Bayesian Assessment of Sample Size for Clinical Trials of Cost-Effectiveness," *Medical Decision Making*, 21, 219–230. [622]

Pocock, S. J., Geller, N. L., and Tsiatis, A. A. (1987), "The Analysis of Multiple End-Points in Clinical-Trials," *Biometrics*, 43, 487–498. [614,616]

Proschan, M. A., and Hunsberger, S. A. (1995), "Designed Extension of Studies Based on Conditional Power," *Biometrics*, 51, 1315–1324. [613]

Spiegelhalter, D., Abrams, K. R., and Myles, J. (2002), *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Chichester: Wiley. [614,616]

Tang, D.-I., Geller, N. L., and Pocock, S. J. (1993), "On the Design and Analysis of Randomized Clinical Trials With Multiple Endpoints," *Biometrics*, 49, 23–30. [614,616]

Tang, D.-I., Gnecco, C., and Geller, N. L. (1989a), "An Approximate Likelihood Ratio Test for a Normal Mean Vector With Nonnegative Components With Application to Clinical Trials," *Biometrika*, 76, 577–583. [614]

——— (1989b), "Design of Group Sequential Clinical Trials With Multiple Endpoints," *Journal of the American Statistical Association*, 84, 776–779. [614]

Westfall, P. H., Krishen, A., and Young, S. S. (1998), "Using Prior Information to Allocate Significance Levels for Multiple Endpoints," *Statistics in Medicine*, 17, 2107–2119. [622]

Zhu, H. J., and Hu, F. F. (2010), "Sequential Monitoring of Response-Adaptive Randomized Clinical Trials," *The Annals of Statistics*, 38, 2218–2241. [613]