Adaptive Group Sequential Three-Arm Trials Including Placebo for Showing Noninferiority of a New Drug

Knapp, Guido
Department of Statistics, TU Dortmund University
44221 Dortmund, Germany
E-mail: guido.knapp@tu-dortmund.de

Hartung, Joachim
Department of Statistics, TU Dortmund University
44221 Dortmund, Germany
E-mail: hartung@statistik.tu-dortmund.de

1. Introduction

The EMEA (1998) clinical trial guideline recommended to include a placebo control group C in a confirmatory phase III trial when an experimental test group T is to be compared with an active reference group R for establishing noninferiority of T with respect to R. Koch (2006) formulated a more detailed regulatory point of view: Essentially, Koch said that "both placebo and active comparator should be included in the trial when the responses of both these treatments cannot be well predicted in the patient population under study. However, it is an ethical mandate that the number of patients randomized to the placebo group should be limited as much as possible. An adaptive design combined with a multiple testing procedure may offer the opportunity to stop recruitment to the placebo group after an interim analysis as soon as superiority of the experimental treatment over placebo has been demonstrated. The trial is then continued into further stages to demonstrate the noninferiority of the experimental treatment in comparison to the reference treatment".

Let $\Delta > 0$ be a noninferiority margin, we test the a-priori ordered hypotheses, that, in short, (I) : $T > C$ and (II) : $T > R - \Delta, \quad \Delta > 0$. When (I) is shown, we can test for (II). For both hypothesis tests, we can take the same significance level which is then the overall significance level in this multiple testing problem. Pigeot et al. (2003) considered a different approach in a one-stage three-arm noninferiority trial. They first required to show $R > C$ before considering other comparisons. This approach bears the risk that the whole study breaks down when R fails to be superior to C although R may represent a very well established treatment.

In this paper, we consider normally distributed response variables with unknown variances in general adaptive group sequential trials, see Hartung (2006). Parameterized p-values, see Cox and Hinkley (1974), of the several stages are combined by use of the inverse normal method well known from meta-analysis, see Hartung, Knapp, and Sinha (2008, Chapter 3). The resulting combined statistics are used for group sequential hierarchical testing of the a priori ordered hypotheses (I) and (II). Further, the concept of repeated confidence intervals, see Jennison and Turnbull (2000) and references cited therein, is extended to the case of unknown variances and possibly adaptively chosen sample sizes in an exact way. Note that, based on the closed testing principle, see Marcus, Peritz, and Gabriel (1976), we can establish the superiority of T over R if the lower bound of the final repeated confidence interval in (II) is positive.

Moreover, we develop formulae for sample size calculation in group sequential trials. These formulae seem to be unknown so far, even in case of non-adaptive group sequential trials, where the computed sample size for the first stage is used in all following stages.

2. Group Sequential Testing

Let us consider a new treatment in a test group T, a standard treatment in a reference group R, and a placebo treatment in a control group C. The associated response variables may be denoted by X_T, X_R, and X_C, which are mutually stochastically independent normally distributed random variables with means μ_T, μ_R, μ_C and variances $\sigma_T^2 > 0$, $\sigma_R^2 > 0$, and $\sigma_C^2 > 0$, respectively, that is,

$$X_T \sim \mathcal{N}\left(\mu_T, \sigma_T^2\right), \quad X_R \sim \mathcal{N}\left(\mu_R, \sigma_R^2\right), \quad X_C \sim \mathcal{N}\left(\mu_C, \sigma_C^2\right).$$

At level $\alpha \in (0, 0.5)$, we first consider the test problem

$$H_0^{TC} : \mu_T \leq \mu_C \quad \text{versus} \quad H_1^{TC} : \mu_T > \mu_C. \tag{1}$$

If H_0^{TC} is rejected in favor of H_1^{TC} at level α, then we consider the family of test problems

$$H_{0,\Delta}^{TR} : \mu_T \leq \mu_R - \Delta \quad \text{versus} \quad H_{1,\Delta}^{TR} : \mu_T > \mu_C - \Delta, \ \Delta \in [0, \Delta_0] \tag{2}$$

at the same level α, where $\Delta_0 \geq 0$ denotes some margin for the noninferiority parameter Δ. This hierarchical testing procedure holds the overall significance level α, see Maurer, Hothorn, and Lehmacher (1995).

We consider a comparative study, which is carried out in a number of independent stages, say K. In the i-th stage, $i = 1, \ldots, K$, let be $\bar{X}_{T_i}$, $\bar{X}_{R_i}$, and $\bar{X}_{C_i}$ the sample means of $n_{T_i} \geq 2$, $n_{R_i} \geq 2$, and $n_{C_i} \geq 2$ responses in the respective treatment groups. The variance parameters can be estimated by the corresponding sample variances $S_{T_i}^2$, $S_{R_i}^2$, and $S_{C_i}^2$, which are stochastically independent of the means and follow scaled χ^2-distributions, that is, for $i = 1, \ldots, K$,

$$(n_{T_i} - 1)\,\frac{S_{T_i}^2}{\sigma_T^2} \sim \chi_{n_{T_i}-1}^2, \quad (n_{R_i} - 1)\,\frac{S_{R_i}^2}{\sigma_R^2} \sim \chi_{n_{R_i}-1}^2, \quad (n_{C_i} - 1)\,\frac{S_{C_i}^2}{\sigma_C^2} \sim \chi_{n_{C_i}-1}^2. \tag{3}$$

The parameters of interest are $\theta_{TC} = \mu_T - \mu_C$ and $\theta_{TR} = \mu_T - \mu_R$. Let t_ν denote the central t-distribution with ν degrees of freedom, then, using Satterthwaite's approximation, it approximately holds in the i-th stage, $i = 1, \ldots, K$,

$$D_i^{TC}(\theta_{TC}) := \frac{\bar{X}_{T_i} - \bar{X}_{C_i} - \theta_{TC}}{\sqrt{S_{T_i}^2/n_{T_i} + S_{C_i}^2/n_{C_i}}} \sim t_{\nu_i(TC)}, \tag{4}$$

$$D_i^{TR}(\theta_{TR}) := \frac{\bar{X}_{T_i} - \bar{X}_{R_i} - \theta_{TR}}{\sqrt{S_{T_i}^2/n_{T_i} + S_{R_i}^2/n_{R_i}}} \sim t_{\nu_i(TR)}, \tag{5}$$

with

$$\nu_i(TC) = \frac{\left(S_{T_i}^2/n_{T_i} + S_{C_i}^2/n_{C_i}\right)^2}{\left(S_{T_i}^2/n_{T_i}\right)^2/(n_{T_i} - 1) + \left(S_{C_i}^2/n_{C_i}\right)^2/(n_{C_i} - 1)},$$

$$\nu_i(TR) = \frac{\left(S_{T_i}^2/n_{T_i} + S_{R_i}^2/n_{R_i}\right)^2}{\left(S_{T_i}^2/n_{T_i}\right)^2/(n_{T_i} - 1) + \left(S_{R_i}^2/n_{R_i}\right)^2/(n_{R_i} - 1)}.$$

Provided $\sigma_T^2 = \sigma_C^2$, $\sigma_T^2 = \sigma_R^2$, or $\sigma_T^2 = \sigma_C^2 = \sigma_R^2$, pooled sample variance estimators and exact t-distributions can be used. We omit the details here.

Let F_{t_ν} denote the cumulative distribution function of a t-variate with ν degrees of freedom, then it holds, for the parameterized $1-p$-values,

$$1-p_i^d(\theta_d) = F_{t_{\nu_i(d)}}\left(D_i^d(\theta_d)\right) \sim \mathcal{U}(0,1), \quad d = TC, TR,\ i = 1,\ldots,K, \tag{6}$$

where $\mathcal{U}(0,1)$ stands for the uniform distribution on the unit interval. Consequently, we obtain

$$z_i^d(\theta_d) := \Phi^{-1}\left(1-p_i^d(\theta_d)\right) \sim \mathcal{N}(0,1), \quad d = TC, TR,\ i = 1,\ldots,K, \tag{7}$$

with Φ^{-1} the inverse of the standard normal cumulative distribution function Φ. Because the stages of the trial are assumed to be independent, we define the combining pivotal quantities

$$Z_j^d(\theta_d) := \sum_{i=1}^{j} z_i^d(\theta_d) \sim \sqrt{j}\ \mathcal{N}(0,1), \quad d = TC, TR,\ j = 1,\ldots,K. \tag{8}$$

Let $Y_1,\ldots Y_K$, in general, be mutually independent $\mathcal{N}(0,1)$-distributed random variables. Then positive critical values $cv_1(d),\ldots,cv_K(d)$ may be defined by the following probability condition:

$$P\left(\sum_{i=1}^{j} Y_i \le cv_j(d) \text{ for all } j = 1,\ldots,K\right) = 1-\alpha, \quad d = TC, TR, \tag{9}$$

for a predefined level $\alpha \in (0, 0.5)$, see Hartung (2006).

Using the critical values $cv_j(d)$ from (9), we get the following probability statements for the combining pivotal quantities (8),

$$P_{\theta_d}\left(Z_j^d(\theta_d) \le cv_j(d) \text{ for } j = 1,\ldots,k \le K\right) \begin{cases} \ge 1-\alpha \text{ for } k < K, \\ = 1-\alpha \text{ for } k = K, \end{cases} \quad d = TC, TR. \tag{10}$$

Consequently, we can formulate the following test procedure at overall level of at most α: We reject H_0^{TC} in favor of H_1^{TC} in (1) at the j-th stage, $1 \le j \le K$, if

$$Z_j^{TC}(0) > cv_j(TC) \quad \text{and} \quad Z_{j^*}^{TC}(0) \le cv_{j^*}(TC),\ j^* = 1,\ldots,j-1. \tag{11}$$

If H_0^{TC} is rejected at the j-th stage, we decide in favor of the alternative $H_{1,\Delta}^{TR}$, $\Delta \in [0, \Delta_0]$, at the k-th stage, $j \le k \le K$, in (2), if

$$\exists k_\Delta \in \{1,\ldots,k\} : Z_{k_\Delta}^{TR}(-\Delta) > cv_{k_\Delta}(TR). \tag{12}$$

If (12) holds, we will stop the trial after the k-th stage. We definitely stop the trial after the K-th stage with either rejection or non-rejection of $H_{0,\Delta}^{TR}$.

3. Group Sequential Confidence Intervals

The functions $F_{t_\nu}(\cdot)$ and $\Phi^{-1}(\cdot)$, used in (6) and (7), are (strictly) monotone increasing in their arguments. The pivotal quantities $D_i^{TC}(\theta_{TC})$ and $D_i^{TR}(\theta_{TR})$ from (4) and (5) are monotone decreasing in θ_{TC} and θ_{TR}, respectively, implying that $z_i^d(\theta_d) = \Phi^{-1}\{F_{t_{\nu_i(d)}}[D_i^d(\theta_d)]\}$, see (7), is monotone decreasing in θ_d, $d = TC, TR$, $i = 1,\ldots,K$. Consequently, $Z_j^{TC}(\theta_{TC})$ and $Z_j^{TR}(\theta_{TR})$ are monotone decreasing in θ_{TC} and θ_{TR}, respectively, $j = 1,\ldots,K$.

Thus, we derive the lower confidence sets on θ_d as

$$CI_{k,\mathrm{I}}^d(\theta_d) := \left\{y \in I\!R \mid Z_j^d(y) \le cv_j(d) \text{ for } j = 1,\ldots,k\right\},\ d = TC, TR,\ k = 1,\ldots,K. \tag{13}$$

The confidence coefficient of $CI_{K,\mathrm{I}}^d$ is at least $1-\alpha$ and exactly $1-\alpha$ for $k = K$, see (10). Further, the confidence sets are nested, that is, $CI_{k+1,\mathrm{I}}^d(\theta_d) \subset CI_{k,\mathrm{I}}^d(\theta_d)$, $k = 1,\ldots,K-1$. Because of the monotonicity in y, the confidence sets are genuine intervals leading to

$$CI^d_{k,\mathrm{I}}(\theta_d) = \left[L^d_k, \infty\right), \quad d = TC, TR, \tag{14}$$

where $L^d_k = \max\{L^d(1), \ldots, L^d(k)\}$ and $L^d(j)$ uniquely solves

$$Z^d_j[L^d(j)] = cv_j(d),\ j = 1, \ldots, k,\ k = 1, \ldots, K. \tag{15}$$

The lower bounds L^d_k from (14) can be now used in the hierarchical testing procedure. In accordance to (11) and (12), we can formulate the following decision rule: We reject H_0^{TC} in favor of H_1^{TC} in (1) at the j-th stage, $1 \le j \le K$, if $L^{TC}_j > 0$ and $L^{TC}_{j^*} \le 0$, $j^* = 1, \ldots, j-1$. If H_0^{TC} is rejected at the j-th stage, we decide in favor of the alternative $H^{TR}_{1,\Delta}$, $\Delta \in [0, \Delta_0]$, at the k-th stage, $j \le k \le K$, in (2), if $L^{TR}_k > -\Delta$. Note if $L^{TR}_k > 0$, we conclude superiority of T over R.

In analogy to (13), let us define the upper confidence sets on θ_d as

$$CI^d_{k,\mathrm{II}}(\theta_d) := \left\{y \in I\!R \mid -cv_j(d) \le Z^d_j(y) \text{ for } j = 1, \ldots, k\right\}, \quad d = TC, TR,\ k = 1, \ldots, K. \tag{16}$$

Again, using (10), each confidence set has a confidence coefficient of at least $1-\alpha$, being exactly $1-\alpha$ for $k = K$. The interval representation is given by

$$CI^d_{k,\mathrm{II}}(\theta_d) = \left(-\infty, U^d_k\right], \quad d = TC, TR, \tag{17}$$

where $U^d_k = \min\{U^d(1), \ldots, U^d(k)\}$ and $U^d(j)$ solves uniquely

$$Z^d_j[U^d(j)] = -cv_j(d),\ j = 1, \ldots, k,\ k = 1, \ldots, K. \tag{18}$$

The two-sided confidence interval on θ_d at stage k is then defined as the intersection of the two corresponding one-sided confidence intervals,

$$CI^d_k(\theta_d) := \left[L^d_k, U^d_k\right], \quad d = TC, TR, \tag{19}$$

where L^d_k is from (14) and U^d_k is from (17), $k = 1, \ldots, K$. The confidence intervals are nested, that is, $CI^d_{k+1}(\theta_d) \subset CI^d_k(\theta_d)$, $k = 1, \ldots, K-1$, $d = TC, TR$, and each confidence interval has a confidence coefficient of at least $1 - 2\alpha$, $0 < \alpha < 1/2$.

Denote $I^d_k(\theta_d) = [L^d(j), U^d(j)]$, see (15) and (18), the individual two-sided confidence interval on θ_d at the k-stage. Then it holds,

$$CI^d_1(\theta_d) = I^d_1(\theta_d) \quad \text{and} \quad CI^d_k(\theta_d) = CI^d_{k-1}(\theta_d) \cap I^d_k(\theta_d), \quad k = 2, \ldots, K,\ d = TC, TR. \tag{20}$$

Since $CI^d_k \subset I^d_k$, the interval I^d_k is another two-sided confidence interval with confidence coefficient of at least $1 - 2\alpha$. The interval I^d_k results from the boundaries in stage k alone and will be always nonempty. Therefore, I^d_k may be preferred to CI^d_k, see for instance Jennison and Turnbull (2000, p. 192) in their corresponding setting. Depending on the choice of α, the two-sided confidence interval $CI^d_k(\theta_d)$ from (19) may be empty. But the probability to obtain an empty confidence interval $CI^d_k(\theta_d)$ is bounded by 2α, $d = TC, TR$.

Instead of the implicitly defined confidence intervals, we provide approximative confidence intervals in an explicit form. Their boundaries may be used as starting points in an iterative procedure to determine the exact confidence intervals.

Let us approximate the central t-distributions involved in the combining pivotal quantities by normal distributions with the same first two moments. The variance of a t_ν-variate is $\nu/(\nu-2)$. So we may define the following weights at the i-th stage, $i = 1, \ldots, K$,

$$w^{TC}_i := \sqrt{\frac{\nu_i(TC) - 2}{\nu_i(TC)[S^2_{T_i}/n_{T_i} + S^2_{C_i}/n_{C_i}]}}, \tag{21}$$

provided $\nu_i(TC) > 2$. Thus, the pivotal quantity $z^{TC}_i(\theta_{TC})$ from (7) is approximated by

$$z^{TC}_i(\theta_{TC})_{appr} = \Phi^{-1}\left(\Phi\left[w^{TC}_i(\bar{X}_{T_i} - \bar{X}_{C_i} - \theta_{TC})\right]\right), \tag{22}$$

which is approximately $\mathcal{N}(0,1)$-distributed. Hence, the combining pivotal quantity $Z_j^{TC}(\theta_{TC})$ from (8) is approximated by

$$Z_j^{TC}(\theta_{TC})_{appr} = \sum_{i=1}^{j} w_i^{TC}(\bar{X}_{T_i} - \bar{X}_{C_i} - \theta_{TC}), \quad j = 1, \ldots, K, \tag{23}$$

which is approximately $\mathcal{N}(0,j)$-distributed. Equating $Z_j^{TC}(y)_{appr}$ to $cv_j(TC)$ and to $-cv_j(TC)$ and solving for y yields the following approximate individual confidence interval on θ_{TC} for $j = 1, \ldots, K$,

$$I_j^{TC}(\theta_{TC})_{appr} = \sum_{i=1}^{j} \frac{w_i^{TC}(\bar{X}_{T_i} - \bar{X}_{C_i})}{\sum_{h=1}^{j} w_h^{TC}} \pm \frac{cv_j(TC)}{\sum_{h=1}^{j} w_h^{TC}}. \tag{24}$$

By setting

$$CI_1^{TC}{}_{appr} = I_1^{TC}(\theta_{TC})_{appr} \quad \text{and} \quad CI_k^{TC}(\theta_{TC})_{appr} = CI_{k-1}^{TC}(\theta_{TC})_{appr} \cap I_k^{TC}(\theta_{TC})_{appr}, \tag{25}$$

$k = 2, \ldots, K$, we obtain approximations of the confidence intervals CI_k^{TC} on $\theta_{TC} = \mu_T - \mu_C$ in (19). Proceeding analogously, we get approximate confidence intervals on $\theta_{TR} = \mu_T - \mu_R$.

4. Group Sequential Point Estimation

For ease of presentation, we describe the group sequential estimation of $\theta_{TC} = \mu_T - \mu_C$. Estimation of $\theta_{TR} = \mu_T - \mu_R$ follows by analogue considerations.

Recall from (8) that the combining statistic $Z_j^{TC}(\theta_{TC})$ is $\mathcal{N}(0,j)$-distributed with mode and median 0. The *maximum likelihood (ML) estimator* $\hat{\theta}_{TC}^{(1)}(j)$ of θ_{TC} at stage j is given by

$$\hat{\theta}_{TC}^{(1)}(j) \text{ solves } Z_j^{TC}\left(\hat{\theta}_{TC}^{(1)}(j)\right) = 0, \quad j = 1, \ldots, K. \tag{26}$$

The solution in (26) is unique.

The *global* p-value at stage j is

$$p_{TC}(j) = 1 - \Phi\left(Z_j^{TC}(\theta_{TC})/\sqrt{j}\right), \quad j = 1, \ldots, K, \tag{27}$$

and solving (27) for θ_{TC} such that $p_{TC}(j) = 1/2$ yields $\hat{\theta}_{TC}^{(1)}(j)$ as solution. Since $Z_j^{TC}(\theta)$ is monotone in θ_{TC}, we can conclude:

$$\hat{\theta}_{TC}^{(1)}(j) \text{ is } \textit{median unbiased}, \quad j = 1, \ldots, K, \tag{28}$$

see Cox and Hinkley (1974, p. 273), that is, the ML-estimator $\hat{\theta}_{TC}^{(1)}(j)$ lies with equal probability as well below the parameter θ_{TC} as above θ_{TC}.

Equating the approximative combining statistic $Z_j^{TC}(\theta_{TC})_{\text{appr}}$ from (23) to 0 and solving for θ_{TC} yields the midpoint of the approximative individual confidence interval $I_j^{TC}(\theta_{TC})_{\text{appr}}$ from (24) as *approximate* median unbiased ML-estimator $\hat{\theta}_{TC}^{(2)}(j)$ of θ_{TC} at the j-th stage, given by

$$\hat{\theta}_{TC}^{(2)}(j) = \sum_{i=1}^{j} \frac{w_i^{TC}(\bar{X}_{T_i} - \bar{X}_{C_i})}{\sum_{h=1}^{j} w_h^{TC}}, \quad j = 1, \ldots, K, \tag{29}$$

where the weights are defined in (21). Note that, in combining the mean differences of the stages, their inverse estimated standard errors are used in the weights and not their inverse estimated variances as known from the 'minimum variance unbiased' estimator of the overall mean difference in meta-analysis, see Hartung, Knapp, and Sinha (2008, Chapter 8). Weighted mean differences like $\hat{\theta}_{TC}^{(2)}(j)$ from (29) are used in the generalized Cochran-Wald statistics considered by Hartung, Böckenhoff, and Knapp (2003).

Replacing in (29) the weights w_i^{TC} by

$$\tilde{w}_i^{TC} = \left(\frac{S_{T_i}^2}{n_{T_i}} + \frac{S_{C_i}^2}{n_{C_i}} \right)^{-1}, \quad i = 1, \ldots, K, \tag{30}$$

we obtain the *meta-analytical* estimator $\hat{\theta}_{TC}^{(3)}(j)$ of θ_{TC} up to the j-th stage, $j = 1, \ldots, K$. For $\theta_{TR} = \mu_T - \mu_R$, the estimators $\hat{\theta}_{TR}^{(h)}(j)$ of θ_{TR} at stage j, $h = 1, 2, 3$, are defined analogously.

5. Sample Size Calculation and Adaptive Updating

Suppressing the subscript i and supposing known variances, let us consider the test statistic

$$D_0^{TR}(\theta_{TR}) = \frac{\bar{X}_T - \bar{X}_R - \theta_{TR}}{\sqrt{\sigma_T^2/n_T + \sigma_R^2/n_R}} \sim \mathcal{N}(0,1), \tag{31}$$

which should be used for testing the point hypotheses $H_0^* : \theta_{TR} = -\Delta$ versus $H_1^* : \theta_{TR} = \theta_{TR}^* > -\Delta$ with fixed $\Delta \in [0, \Delta_0]$ and fixed value $\theta_{TR}^* > -\Delta$. So, $D_0^{TR}(-\Delta) \sim \mathcal{N}(0,1)$ under H_0^*. Given level $\alpha \in (0,1)$ and desired power $1 - \beta_{TR}$, $\beta_{TR} \in (0,1)$, the required sample sizes n_T and n_R should satisfy

$$\frac{\theta_{TR}^* - (-\Delta)}{\sqrt{\sigma_T^2/n_T + \sigma_R^2/n_R}} \geq \Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta_{TR}). \tag{32}$$

Let stage 0 denote a-priori information and external restrictions. After stage j, let $\hat{\theta}_{TR}(j) > -\Delta$, $\widehat{\sigma_T^2}(j)$, and $\widehat{\sigma_R^2}(j)$, $j = 0, 1, \ldots, K-1$, be reasonable estimates of their corresponding parameters based on previous information of stages $0, 1, \ldots, j$. Consider the above test of the point hypotheses and replace the unknown parameters by their estimates in (32), we obtain

$$\frac{\hat{\theta}_{TR}(j) + \Delta}{\sqrt{\widehat{\sigma_T^2}(j)/n_T + \widehat{\sigma_R^2}(j)/n_R}} \geq \Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta_{TR}), \quad j = 0, \ldots, K-1. \tag{33}$$

Note that $\hat{\theta}_{TR}(j) + \Delta > 0$ must be fulfilled.

By analogue considerations, with $1 - \beta_{TC}$, $\beta_{TC} \in (0,1)$, the desired power at $\theta_{TC} = \hat{\theta}_{TC}(j) > 0$ in the test problem (1), the required sample sizes n_T and n_C after stage j should satisfy

$$\frac{\hat{\theta}_{TC}(j)}{\sqrt{\widehat{\sigma_T^2}(j)/n_T + \widehat{\sigma_C^2}(j)/n_C}} \geq \Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta_{TC}), \quad j = 0, \ldots, K-1, \tag{34}$$

where $\hat{\theta}_{TC}(j)$ and $\widehat{\sigma_C^2}(j)$ are reasonable estimates of their corresponding parameters based on previous information of stages $0, 1, \ldots, j$.

Let us define the sets of feasible sample sizes, $k = 0, \ldots, K-1$,

$$\Gamma_{TR}(\kappa, \beta_{TR}, \Delta)_k := \{(n_T, n_R) \in I\!N \times I\!N | n_T \text{ and } n_R \text{ satisfy (33) for } j = k \text{ and } \alpha = \kappa\} \tag{35}$$

$$\Gamma_{TC}(\kappa, \beta_{TC})_k := \{(n_T, n_C) \in I\!N \times I\!N | n_T \text{ and } n_C \text{ satisfy (34) for } j = k \text{ and } \alpha = \kappa\}. \tag{36}$$

For $d = TC, TR$, recall from (9) the event

$$\left\{ \sum_{i=1}^{h} Y_i \leq cv_h(d) \text{ for all } h = 1, \ldots, K \right\},$$

and let us consider for an arbitrary, but fixed, stage j, $j \in \{1, \ldots, K\}$, the event

$$\left\{ \sum_{i=1}^{h} Y_i \leq cv_h \text{ for } h = 1, \ldots, j-1, \text{ and } \sum_{i=1}^{j-1} Y_i + \sum_{i=j}^{K} Y_i \leq cv_K(d) \right\}.$$

Note that $\sum_{i=j}^{K} Y_i$ is $\mathcal{N}(0,\ K-(j-1))$-distributed and may be collapsed to $\sqrt{K-(j-1)}Y_j$ having the same distribution. Hence, we obtain

$$P\left\{\sum_{i=1}^{h} Y_i \le cv_h(d) \text{ for } h=1,\ldots,j-1, \text{ and } \sum_{i=1}^{j-1} Y_i + \sqrt{K-(j-1)}Y_j \le cv_K(d)\right\}$$

$$\ge P\left\{\sum_{i=1}^{h} Y_i \le cv_h(d) \text{ for all } h=1,\ldots,K\right\}. \tag{37}$$

Let θ_d^0 denote a value for θ_d under the null-hypothesis H_0^d, given as H_0^{TC} from (1) for $d=TC$ or as $H_{0,\Delta}^{TR}$ from (2) for $d=TR$ and $\Delta \in [0,\Delta_0]$ fixed. Let us assume we decide after stage $j-1$ to omit the interim analyses j up to $K-1$. Then, we can assign the remaining weight $\sqrt{K-(j-1)}$ to the next final study part, named stage (j,K), and build the final test statistic,

$$Z_{(j,K)}^d(\theta_d^0) = Z_{j-1}^d(\theta_d^0) + \sqrt{K-(j-1)}\ \Phi^{-1}\left[1-p_{(j,K)}^d(\theta_d^0)\right], \tag{38}$$

where $Z_{(j,K)}^d(\theta_d^0) \sim \sqrt{K}\ \mathcal{N}(0,1)$ under $H_0^d,\ j=1,\ldots,K$, defining $Z_0^d=0$. We would then reject H_0^d if $Z_{(j,K)}^d(\theta_d^0) > cv_K(d)$, at level of at most α by (9) and (37).

Equating $Z_{(j,K)}^d(\theta_d^0)$ from (38) to $cv_K(d)$ and solving for $p_{(j,K)}^d(\theta_d^0)$ yields the projected p-value

$$\hat{p}_{(j,K)}^d(\theta_d^0) = 1-\Phi\left[\frac{cv_K(d)-Z_{j-1}^d(\theta_d^0)}{\sqrt{K-(j-1)}}\right], d=TC,TR, j=1,\ldots,K, \tag{39}$$

which is the probability of false rejection of the true null hypothesis in one next and final step given the results of the first $j-1$ stages. Thus, the projected p-value can be regarded as a conditional error function. Consequently, we plan the final stage (j,K) at level

$$\alpha_{(j,K)}^d = \hat{p}_{(j,K)}^d(\theta_d^0), d=TC,TR,\ j=1,\ldots,K. \tag{40}$$

Conditioned on $\theta_{TC}=\hat{\theta}_{TC}(j-1)>0$ and $\theta_{TR}=\hat{\theta}_{TR}(j-1)>-\Delta$, the required sample sizes M_{T_j}, M_{C_j}, and M_{R_j} of the respective groups in the final stage (j,K), attaining power $1-\beta_{TC}$ for $d=TC$ in (1) and power $1-\beta_{TR}$ for $d=TR$ in (2), should be feasible and satisfy:

$$(M_{T_j}, M_{C_j}) \in \Gamma_{TC}\left(\hat{p}_{(j,K)}^{TC}(0), \beta_{TC}\right)_{j-1} \qquad \text{and} \tag{41}$$

$$(M_{T_j}, M_{R_j}) \in \Gamma_{TR}\left(\hat{p}_{(j,K)}^{TR}(-\Delta), \beta_{TR}, \Delta\right)_{j-1}, \tag{42}$$

see (35), (36), (39), (40).

If we do not want to finish the trial in this way and have in mind the originally planned $K-(j-1)$ further stages, we will choose the sample size in each group for stage j proportionally as

$$n_{T_j} \approx \frac{M_{T_j}}{K-j+1}, \quad n_{C_j} \approx \frac{M_{C_j}}{K-j+1}, \quad n_{R_j} \approx \frac{M_{R_j}}{K-j+1}, \quad j=1,\ldots,K. \tag{43}$$

Note that each sample size should be at least 2 in each stage.

Especially for $j=1$:

$$n_{T_1} \approx \frac{M_{T_1}}{K},\ n_{C_1} \approx \frac{M_{C_1}}{K},\ \text{and } n_{R_1} \approx \frac{M_{R_1}}{K} \tag{44}$$

where, see (35) and (36),

$$(M_{T_1}, M_{C_1}) \in \Gamma_{TC}\left(\alpha_{TC}, \beta_{TC}\right)_0, \quad \alpha_{TC} := 1-\Phi(cv_K(TC)/\sqrt{K}),$$

$$(M_{T_1}, M_{R_1}) \in \Gamma_{TR}\left(\alpha_{TR}, \beta_{TR}, \Delta\right)_0, \quad \alpha_{TR} := 1-\Phi(cv_K(TR)/\sqrt{K}),$$

are feasible *starting sample sizes*.

Taking the initial sample sizes from (44) in all stages, we obtain *formulae for sample size calculation* in non-adaptive group sequential trials.

We start with the above calculated initial sample sizes in the first stage of the study. Then, using the above procedure, we reach the full power $1-\beta_{TC}$, conditioned on $\theta_{TC}=\hat{\theta}_{TC}(K-1)>0$, and $1-\beta_{TR}$, conditioned on $\theta_{TR}=\hat{\theta}_{TR}(K-1)>-\Delta$, latest in stage $j=K$. The *total power*, say $1-\beta_{\text{Total}}$, of the hierarchical testing of (1) and (2), is then bounded by

$$1-\beta_{TC}-\beta_{TR}\leq 1-\beta_{\text{Total}}\leq \min\{1-\beta_{TC},1-\beta_{TR}\}. \tag{45}$$

Further, we may formally define the p-values, see (6), as suiting to the null-hypothesis that θ_d is the *true* parameter, see Cox and Hinkley (1974, p. 221). So, we may apply the general result that under the null-hypothesis p-values preserve their distribution and independence (for continuous null-distributions) when sample sizes are chosen adaptively in a consecutive way, see for instance Brannath, Posch, and Bauer (2002). All the above procedures are based on such p-values. Consequently, all the statements remain valid when sample sizes are chosen adaptively as demonstrated in this section, see also Hartung (2006).

6. Final remarks

In this paper, we have introduced an adaptive group-sequential analysis for a three-arm trial including placebo for showing noninferiority of a new drug. In the talk, we discuss an example to show the practical implication of this procedure. Slides of the talk are available from the authors upon request.

In Section 2, we have defined positive one-sided critical values $cv_j,\ j=1,\ldots,K$, by the probability condition (9). For a fixed number of stages K and an overall significance level α, we get an O'Brien and Fleming (1979) design with constant critical values in (9), say $cv_j=cons_{OBF}(K,\alpha)$, and a Pocock (1977) design with monotone increasing critical values given as $cv_j=\sqrt{j}\ cons_{PO}(K,\alpha), j=1,\ldots,K$, see Hartung (2006), where also some of these one-sided critical values are tabulated. Designs with intermediate values of the critical values are considered, for instance, in Jennison and Turnbull (2000). Usually, two-sided critical values at level 2α for the corresponding symmetric two-sided tests are tabulated in literature. For $K\geq 2$, these two-sided critical values are slightly smaller than the one-sided critical values at level α. At least for $\alpha\leq 0.05$, these two-sided critical values may be used here for practical applications, see Jennison and Turnbull (2000, p. 192).

We have defined the two-sided confidence interval CI_k, see (19), as the intersection of the one-sided intervals $CI_{k,I}$ and $CI_{k,II}$, see (14) and (17), and the confidence coefficient of CI_k is at least $1-2\alpha$. If we use the critical values of the correspondent two-sided tests at level 2α, we get a two-sided confidence interval, say CI_k^0, that is slightly narrower than CI_k for $K\geq 2$, but has a confidence coefficient of at least $1-2\alpha$ as well. Moreover, the final CI_K^0 reaches a confidence coefficient of exactly $1-2\alpha$.

In Section 5, we have computed sample sizes n using a normal approximation for applying t-variates. Nearly exact values are achieved by correcting the sample size n with the variance of a t_{n-1}-variate, that is, replacing n by $n_{\text{corr}}=n(n-1)/(n-3),\ n\geq 4$. The idea behind the correction is the same as in replacing a t-variate by a normal variate with identical variance. However, computed values have usually to be modified to fit some side conditions like block randomization schemes.

REFERENCES

Brannath, W., Posch, M., Bauer, P. (2002). Recursive combination tests. *JASA* 97:236–244.

Cox, D.R., Hinkley, D.V. (1974). *Theoretical Statistics*. New York: Chapman and Hall.

EMEA (The European Agency for the Evaluation of Medicinal Products) (1998). ICH Topic E9: Statistical Principles for Clinical Trials, London, CPMP/ICH/363/96.

Hartung, J. (2006). Flexible designs by adaptive plans of generalized Pocock- and O'Brien-Fleming-type and by Self-designing clinical trials. *Biometrical J.* 48:521–536.

Hartung, J., Böckenhoff, A., Knapp, G. (2003). Generalized Cochran-Wald statistics in combining of experiments. *J. Statist. Plann. Inference* 113:215–237.

Hartung, J., Knapp, G., Sinha, B.K. (2008). *Statistical Meta-Analysis with Applications*. New York: Wiley.

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

Koch, A. (2006). Confirmatory clinical trials with an adaptive design. *Biometrical J.* 48:574–585.

Marcus, R., Peritz, E., Gabriel, K.R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63:655–660.

Maurer, W., Hothorn, A., Lehmacher, W. (1995). Multiple comparisons in drug clinical trials and preclinical assays: A-priori ordered hypotheses. In: Vollmer, J. (ed.) (1995). *Testing Principles in Clinical and Preclinical Trials*. New York: Gustav Fischer.

O'Brien, P.C., Fleming, T.R. (1979). A multiple testing procedure for clinical trials. *Biometrics* 35:549–556.

Pigeot, T., Schäfer, J., Röhmel, J., Hauschke, D. (2003). Assessing non-inferiority of a new drug in a three-arm clinical trial including a placebo. *Statist. Med.* 22:883–399.

Pocock, S.J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* 64:191–199.