Adaptive designs with arbitrary dependence structure

Rene Schmidt, Andreas Faldum, Joachim Gerß

Institute of Biostatistics and Clinical Research, University of Münster

rene.schmidt@ukmuenster.de

Introduction

Interim analyses are used in clinical trials to control the trial progress, since the investigator is interested in making an early decision due to ethical and economic reasons. Whenever repeated significance testing is performed, it has to be assured that no inflation of type I error rate occurs. In classical adaptive designs like the design proposed by Bauer (1989) and Bauer and Köhne (1994), stagewise test statistics are calculated with data collected during the different stages. If the test statistics are independent or if the weaker p-clud condition (Brannath, Posch and Bauer, 2002) is fulfilled, the adaptive design preserves the significance level. A possibility to achieve independent test statistics is to exploit the independent increment structure of many statistical models (Jennison and Turnbull, 2000).

Recent trials have suggested the need for approaches, where several outcomes are viewed as key endpoints while controlling the type I error rate, rather than selecting just one of these outcomes as primary endpoint and treating the remaining outcomes as secondary. An adaptive generalization of the closed testing procedure of Marcus, Peritz and Gabriel (1976) was proposed by Hommel (2001). It turns out that adaptive closed testing procedures can result in adaptive designs with dependent p-values if an adaptive modification of hypotheses is performed. If the dependence structure is not taken into account adequately, control of type I error rate is endangered. Therefore there is a need to consider adaptive designs with dependent p-values.

For two-stage adaptive designs with correlated test statistics, Hommel, Lindig and Faldum (2005) proposed the modified Simes test (MST) and Götte, Hommel and Faldum (2009) presented a general framework, which applies if the test statistics underlying the two stages follow a bivariate normal distribution. We are concerned with two-stage adaptive designs which guarantee control of the type I error rate irrespective of the true dependence structure between the p-values P_1 and P_2 of the stages (hereafter referred to as worst case designs), assuming that $P_{H_0}(P_i \le p_i) \le p_i,\ i = 1{,}2$. The simplest adaptive design in this setting is the Bonferroni design, which might be defined by the rejection region

$$\{P_1 \le \alpha_1\} \cup \{\alpha_1 < P_1 \le \alpha_0, P_2 \le \alpha^*\}. \qquad (1)$$

Here, α_1 denotes the significance level of the first stage, α_0 is the stop for futility bound, and $0 \le \alpha^* \le 1$ is a fixed constant independent of P_1. Then the type I error rate $\alpha_1 + \alpha^*$ is controlled irrespective of the dependence structure of the p-values. It is a natural question to investigate the performance of adaptive designs with arbitrary conditional error function $\alpha(P_1)$ in the presence of dependent p-values.

Worst case inverse normal designs

Consider two-stage adaptive designs without futility stop and equally weighted stages with conditional error function of the form:

$$\alpha(p_1) = \begin{cases} 1 & ,\text{if } p_1 \le \alpha_1 \\ 1 - F(\sqrt{2}c - F^{-1}(1-p_1)) & ,\text{if } p_1 > \alpha_1 \end{cases} \qquad (2)$$

where $\mathrm{F}: \mathrm{R} \to [0{,}1]$ is any strictly increasing continuous function. With F chosen as the standard normal distribution function Φ, we obtain a design of inverse normal type (Lehmacher and Wassmer, 1999). The type I error rate $P_{H_0}(P_1 \le \alpha_1) + P_{H_0}(P_2 \le \alpha(P_1), P_1 > \alpha_1)$ for this design is

$$1 - P_{H_0}(Z_1 + Z_2 < \sqrt{2}c, Z_1 < \gamma) \qquad (3)$$

with $Z_i := \Phi^{-1}(1-P_i)$ for $i = 1{,}2$, and $\gamma := \Phi^{-1}(1-\alpha_1)$. In order to control the type I error rate α irrespective of the true joint distribution of P_1 and P_2, the design parameter c of the two-stage adaptive design (2) with $\mathrm{F} = \Phi$ has to be chosen such that

$$\alpha = 1 - \inf_{C} P_{H_0}(Z_1 + Z_2 < \sqrt{2}c, Z_1 < \gamma) \qquad (4)$$

where the infimum is over all copulas C of Z_1 and Z_2 (worst case). This infimum may be computed in analytically closed form and (4) may be solved for the design parameter c of the adaptive design.

Theorem: Let $0 \le \alpha_1 \le \alpha \le ½$ and c be the design parameters of the two-stage inverse normal design corresponding to the conditional error function (2) with $\mathrm{F} = \Phi$. Assume that the p-values P_1 and P_2 of the stages are stochastically larger than the uniform distribution on $[0{,}1]$ under the null hypothesis.

Then a type I error rate α is controlled in the worst case (i.e. irrespective of the dependence structure of P_1 and P_2), if the design parameter c is chosen as follows in terms of α and α_1:

- If $2\alpha_1 \le \alpha$, then $c = \sqrt{2}\,\Phi^{-1}(1 - \frac{\alpha}{2})$.
- If $2\alpha_1 \le \alpha$, then $c = \{\Phi^{-1}(1-\alpha_1) + \Phi^{-1}(1+\alpha_1-\alpha)\}/\sqrt{2}$.

If the p-values P_1 and P_2 are $[0{,}1]$ uniform random variables and c is chosen as above, there is a copula C of P_1 and P_2 such that the type I error rate of the adaptive design (2) with $\mathrm{F} = \Phi$ equals α.

Comparison with the reference level for independent and uniformly distributed p-values

Under the assumption of independent and uniformly distributed p-values, the type I error rate α_{ind} for the two-stage inverse normal design (2) with $\mathrm{F} = \Phi$ is

$$\alpha_{ind} = 1 - \int_{-\infty}^{\gamma} \Phi(\sqrt{2}c - z_1)\phi(z_1)dz_1 . \qquad (5)$$

If the design is chosen from the Δ-family of Wang and Tsiatis (1987), c and γ are related by $c = \gamma \cdot 2^{\Delta - 0.5}$. In Figure 1, the type I error rate α in the worst case is displayed for the two-stage adaptive design (2) with $\mathrm{F} = \Phi$ from the Δ-family, if c is chosen such that the significance level α_{ind} is exactly preserved for independent and uniformly distributed p-values. Three scenarios are displayed in Figure 1: (i) O'Brien and Fleming design (Δ=0), (ii) Pocock design (Δ=0.5), and (iii) design with Δ =1. Notice that, $c \le \sqrt{2}\gamma$ and thus $2\,\alpha_1 \le \alpha$ for all $\Delta \le 1$.

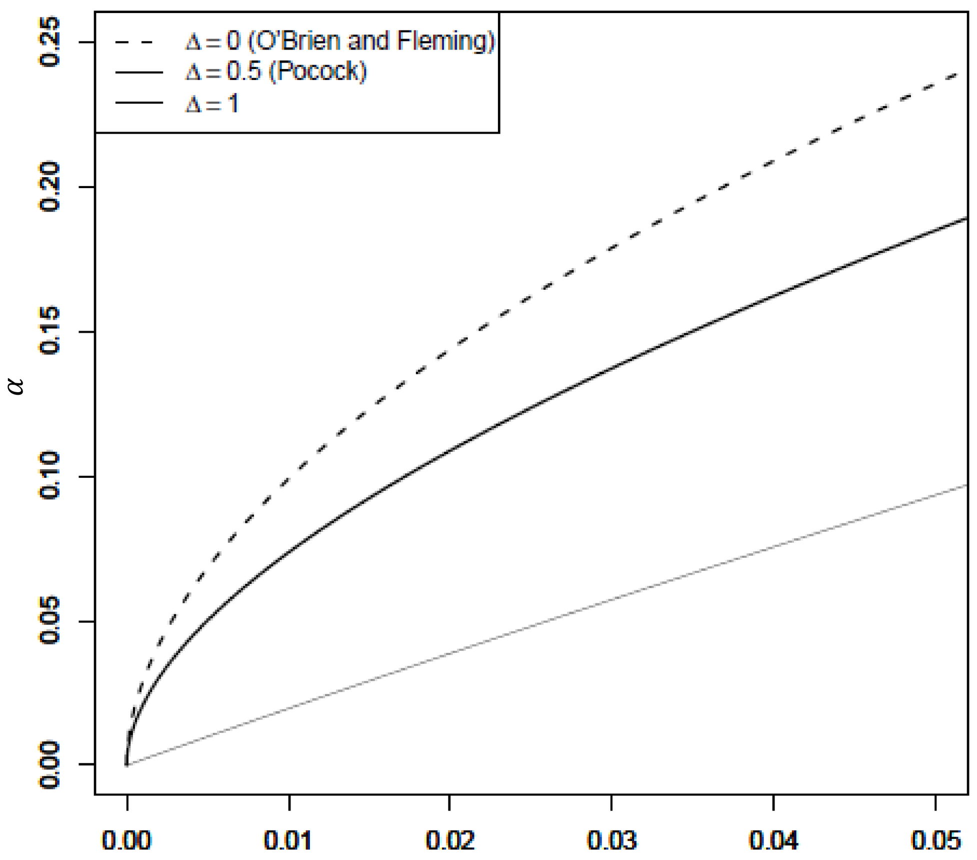

Figure 1.

Worst case Bauer and Köhne designs

With the choice $F(x) = 1 - \exp(-x)$ in (2), we obtain a design of Bauer and Köhne (1994) type: One finds $1 - F\left(\sqrt{2}c - F(1-P_1)\right) = c_{BK}/p_1$ with $c_{BK} = \exp(-\sqrt{2}c)$ denoting the parameter of the design of Bauer and Köhne (1994). Again, for any choice of the design parameters α_1 and c_{BK}, the type I error rate α in the worst case may be determined in analytically closed form and may thus be compared with the reference level α_{ind} for independent and uniformly distributed p-values (see Figure 2). The reference level for independent and uniformly distributed p-values is with $f(x) = F'(x) = \exp(-x)$:

$$\alpha_{ind} = 1 + \int_0^{\min(\gamma,\sqrt{2}c)} F(\sqrt{2}c - z_1) f(z_1) dz_1 \qquad (6)$$

Two scenarios are displayed in Figure 2: (i) $\alpha_1 = c_{BK}$, and (ii) $\alpha_1 = \alpha_{ind}/2$.

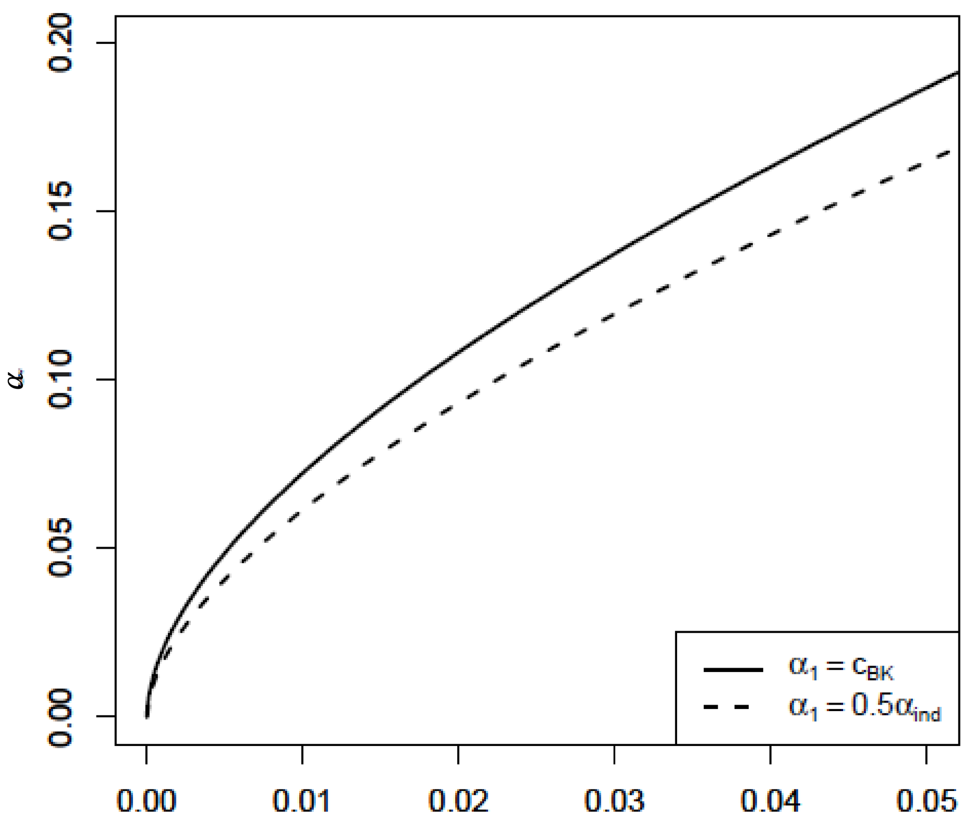

Figure 2.

Discussion

Performance of worst case inverse normal designs and worst case Bauer and Köhne designs without futility stop was studied in order to facilitate a comparison with the Bonferroni design. Amongst the inverse normal designs considered above, the worst case design of O'Brien and Fleming type turned out to be the most conservative design as compared to the reference level for independent and uniformly distributed p-values; followed by the Pocock type worst case design and the worst case design for Δ= 1 . The latter one resulted in the smallest inflation of the type I error rate. However, for increasing Δ, more and more type I error is already spent in the interim analysis (Δ = 1 and $\alpha_{ind} = 2.5\%$ already implies $\alpha_1 = 2.4\%$). For the designs of Bauer and Köhne type studied above, inflation of the type I error rate in the worst case is comparable with that of the above Pocock design. In situations of practical relevance, the Bonferoni design without futility stop is only slightly conservative: For example, choosing $\alpha^* = 1.25\%$, $\alpha_1 = 1.25\%$ and $\alpha_0 = 1$, the Bonferroni design controls a type I error rate of $\alpha = 2.5\%$ in the worst case. For independent und uniformly distributed p-values, a significance level of 2.484% is still attained. Consequently, the Bonferroni design is considerably less conservative than the worst case designs of Pocock or O'Brien and Fleming type without futility stop considered above.

All in all, the Bonferroni design performs best amongst the adaptive designs without futility stop for strongly dependent p-values considered above. Thereby, performance of adaptive designs was evaluated in terms of the worst case type I error rate as compared to the reference level for independent and uniformly distributed p-values. Further research will be devoted to aspects of sample size and average sample number (ASN) in order to complete assessment of worst case adaptive designs. In the long term, it would be desirable to find a class of conditional error functions, which performs optimal in the setting of strongly dependent p-values.

Bibliography

Bauer, P. (1989). Multistage testing with adaptive design. *Biometrie und Informatik in Medizin und Biologie* **20(4)**, 130-136.

Bauer, P. and Köhne, K. (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics* **50**, 1029-1041.

Götte, H., Hommel, G. and Faldum, A. (2009). Adaptive designs with correlated test statistics. *Statistics in Medicine* **28**, 1429-1444.

Hommel, G. (2001). Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal* **43**, 581-589.

Hommel, G., Lindig, V. and Faldum, A. (2005). Two-stage adaptive designs with correlated test statistics. *Journal of Biopharmaceutical Statistics* **15**, 613-623.

Jennison, C.J. and Turnbull, B.W. (2000). *Group Sequential Methods with Applications to Clinical Trials, London:* Chapman & Hall.

Lehmacher, W. and Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* **55**, 1286-1290.

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

Wang, S. K. and Tsiatis, A. A. (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics* **43**, 193-199.