Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 1956–1961
journal of statistical planning and inference
www.elsevier.com/locate/jspi

On the efficiency of adaptive designs for flexible interim decisions in clinical trials

Werner Brannath*, Peter Bauer, Martin Posch

Department of Medical Statistics, University of Vienna, Schwarzspanierstr. 17, A-1090 Vienna, Austria

Available online 24 August 2005

Abstract

It is shown that the optimal group sequential designs considered in Tsiatis and Mehta [2003. On the inefficiency of the adaptive design for monitoring clinical trials. Biometrika 90, 367–378] are special cases of the more general flexible designs which allow for a valid inference after adapting a predetermined way to spend the rejection and acceptance probabilities. An unforeseen safety issue in a clinical trial, for example, could make a change of the preplanned number of interim analyses and their sample sizes appropriate. We derive flexible designs which have equivalent rejection and acceptance regions if no adaptation is performed, but at the same time allow for an adaptation of the spending functions, and have a conditional optimality property.

Keywords: Adaptive design; α-spending; Group-sequential test; Inefficiency; Likelihood ratio test

1. Introduction

In a recent paper, Tsiatis and Mehta (2003) show that adaptive designs can be uniformly improved by group sequential designs which are based on the sequential likelihood ratio test statistics. This improvement, however, is possible only under the assumption that the appropriate spending functions and sample sizes (power characteristics) can be completely specified a priori and that interim analyses can be performed at no cost. In a specific clinical trial, however, an adaptation of the sample sizes and the error spending functions could be of advantage: usually not only the primary end point is of importance. Other information, e.g.,

* Corresponding author.
E-mail address: werner.brannat@meduniwien.ac.at (W. Brannath).

0378-3758/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2005.08.014

on safety end points, economic factors like treatment costs or costs of an interim analysis play an important role as well.

Examples: (i) An unforeseen safety issue arising at the first interim analysis could imply that the information on the safety parameters at later interim analyses would be insufficient to justify early termination of the trial. In this case it is more efficient to skip the hypothesis tests at later interim analyses (i.e., not to spend any level there) and recalculate the decision boundaries for the final analysis accordingly. (ii) In another example new financial resources (available, e.g. due to a promising interim result on the primary or secondary endpoints) may allow for an increase of the overall or conditional power by enlarging the maximum sample size thereby increasing the chance for a successful trial with more precise effect estimates. (iii) Due to promising results from an early interim analysis and because of increased competition (e.g. another company is starting a similar program) sponsors wish to include an extra interim analysis before the planned final analysis to shorten the development procedure.

Note that the adaptations mentioned in the examples, when based on the unblinded interim data, are not possible within the classical α-spending approach, since they might inflate the overall type I error rate, and hence invalidate the trial results. Therefore, methods which allow for unscheduled design adaptations are useful, even though they can be theoretically improved if all relevant information were available in advance.

In the sequel it is shown how to change the level and power spending characteristic of a preplanned sequential likelihood ratio test using the adaptive design methodology. If no adaptation is performed, the resulting rejection and acceptance regions are equivalent to the sequential likelihood ratio test, and with adaptations the design is optimal, at least, in a conditional sense.

2. Adaptive designs

Adaptive designs have been introduced for a control of type I error probability in cases where information from unblinded interim data are used for adapting the design mid-trial (Bauer, 1989; Bauer and Köhne, 1994; Proschan and Hunsberger, 1995). These designs can be defined in terms of a conditional error function which has to be prespecified for every time point where the trial design may be altered (Proschan and Hunsberger, 1995; Posch and Bauer, 1999; Müller and Schäfer, 2001, 2004). The conditional error function at a specific time point t of the trial is the probability under the null hypothesis H_0 of rejecting H_0 conditional on the data Y_t accumulated so far. More formally, the conditional error function $A_t(Y_t)$ at time t is a deterministic measurable function of Y_t with values in the unit interval $[0, 1]$ such that $E_0[A_t(Y_t)] = \alpha$ where E_0 denotes the expectation under H_0 and α is the level of the trial. At time t one can choose the design for the remainder of the trial (based on all available information) with a conditional rejection probability of at most $A_t(Y_t)$ under H_0. Thus with P_0 the null probability, the new design needs to satisfy

$$P_0(\text{reject } H_0 | Y_t) \leqslant A_t(Y_t), \tag{1}$$

where the right-hand side of (1) is computed from the conditional null distribution of the forthcoming data observed after t (according to the new design) given the data Y_t observed

until t. If adaptations are done in a measurable way, then

$$P_0(\text{reject } H_0) = E_0[P_0(\text{reject } H_0|Y_t)] \leqslant E_0[A_t(Y_t)] = \alpha. \tag{2}$$

Hence, the overall type I error rate is at most α, irrespective of the adaptations.

Remarks: (i) Liu et al. (2002) show how to deduce (2) from (1) if adaptations are restricted to a countable set. They also give an example for an uncountable set of adaptations that satisfy (1) up to null sets but violate (2). Hence, a strict proof of (2) is available only for countable adaptations even when disjoint samples are recruited. However, this is not a severe restriction for the practical use of adaptive designs, since the set of design parameters as well as the set of values for each design parameter will in practice be always finite (even if not prespecified in the protocol). (ii) It can become impossible to compute the conditional error $A(Y_t)$ of the original design and the conditional rejection probability $P_0(\text{reject } H_0|Y_t)$ of the new design given the complete interim data Y_t. Such a situation appears if nuisance parameters are involved, for instance, in the context of survival studies if patients recruited before t have events after t and the interim data contain information on covariables or safety endpoints of these patients (cf. Schäfer and Müller, 2001; Bauer and Posch, 2004). To compute conditional rejection probabilities given the data Y_t would require the knowledge of the joint distribution of all covariables (potentially driving the adaptations) and the endpoint for H_0. This distribution depends on nuisance parameters that are usually unknown. (iii) If disjoint cohorts of patients are recruited before and after t then the conditional error function $A(Y_t)$ of the original design and the conditional type I error probability $P_0(\text{reject } H_0|Y_t)$ of the new design are independent from the information of covariables of patients recruited before t.

3. Turning an optimally preplanned design into a flexible design

It has been argued by Müller and Schäfer (2001) that using the conditional error function of a group sequential design allows one to combine optimality properties of group sequential designs with the flexibility of adaptive designs. This will be illustrated for the sequential likelihood ratio test in Tsiatis and Mehta (2003).

3.1. Sequential likelihood ratio test

A group sequential trial (whether adaptive or not) is represented in Tsiatis and Mehta (2003) by a specific number of stages $1, \ldots, K$. The data obtained at stage j is represented by the random vector X_j with values in a finite dimensional real vector space, and the cumulative data of stage j by the random vector $Y_j = (X_1, \ldots, X_j)$. We assume as in Tsiatis and Mehta (2003) that the null and alternative hypotheses H_0 and H_1 are simple and imply continuous and positive densities $p_{0j}(x_j)$ and $p_{1j}(x_j)$ for X_j. We further assume that X_i and X_j are independent for $i \neq j$ under H_0 and under H_1. A sequential test based on such data is a sequence of Y_j-measurable rejection and acceptance regions $(\mathscr{R}_j, \mathscr{A}_j)$ and the trial is stopped at stage j if either $Y_j \in \mathscr{R}_j$ (rejection) or $Y_j \in \mathscr{A}_j$ (acceptance). We let further $\bar{\mathscr{R}}_j = \mathscr{R}_1 \cup \cdots \cup \mathscr{R}_j$ and $\bar{\mathscr{A}}_j = \mathscr{A}_1 \cup \cdots \cup \mathscr{A}_j$, and define the probabilities $\alpha_j = P_0(\bar{\mathscr{R}}_j)$ and $\theta_j = P_0(\bar{\mathscr{A}}_j)$ under the null hypothesis H_0. The sequence $\alpha_1 \leqslant \cdots \leqslant \alpha_K$ corresponds to the

α-spending function of Lan and DeMets (1983), and the sequence $\theta_1 \leqslant \cdots \leqslant \theta_K$ is denoted by the θ-spending sequence. Spending sequences of group sequential trials necessarily satisfy $\theta_j + \alpha_j < 1$ for all $j < K$ and $\theta_K + \alpha_K = 1$. Note that α_K equals the overall level α.

Tsiatis and Mehta (2003) show that given the α- and θ-spending sequence the test based on the likelihood ratios is optimal in the sense that the rejection probability under H_1 and the acceptance probability under H_0 are maximal for all stages j. Let $L_j = p_{1j}(X_j)/p_{0j}(X_j)$ be the likelihood ratio of the data obtained at stage j and denote $\bar{L}_j = \Pi_{i=1}^{j} L_i$ $(j = 1, \ldots, K)$ the sequential likelihood ratio test statistics. Given any α- and θ-spending sequence one can define a group-sequential test which is based on the sequential likelihood ratio test statistics. To this end we let the first stage rejection and acceptance regions be $\mathscr{R}_1^{\mathrm{LR}} = \{\bar{L}_1 > b_1\}$ and $\mathscr{A}_1^{\mathrm{LR}} = \{\bar{L}_1 < a_1\}$ with a_1 and b_1 such that $P_0(\mathscr{R}_1^{\mathrm{LR}}) = \alpha_1$ and $P_0(\mathscr{A}_1^{\mathrm{LR}}) = \theta_1$. For the stages $j = 2, \ldots, K$ we let $\mathscr{R}_j^{\mathrm{LR}} = \{a_1 \leqslant \bar{L}_1 \leqslant b_1, \ \ldots, a_{j-1} \leqslant \bar{L}_{j-1} \leqslant b_{j-1}, \bar{L}_j > b_j\}$ and $\mathscr{A}_j^{\mathrm{LR}} = \{a_1 \leqslant \bar{L}_1 \leqslant b_1, \ldots, a_{j-1} \leqslant \bar{L}_{j-1} \leqslant b_{j-1}, \bar{L}_j < a_j\}$ with b_j and a_j such that $P(\mathscr{R}_j^{\mathrm{LR}}) = \alpha_j - \alpha_{j-1}$ and $P(\mathscr{A}_j^{\mathrm{LR}}) = \theta_j - \theta_{j-1}$. Note that from $\theta_K + \alpha_K = 1$ we get $a_K = b_K$. According to Theorem 1 in Tsiatis and Mehta (2003) we have $P_1(\bar{\mathscr{R}}_j) \leqslant P_1(\bar{\mathscr{R}}_j^{\mathrm{LR}})$ and $P_0(\bar{\mathscr{A}}_j) \leqslant P_0(\bar{\mathscr{A}}_j^{\mathrm{LR}})$ for all stages j and for every sequential test $(\mathscr{R}_j, \mathscr{A}_j)$, $j = 1, \ldots, K$, which has the same α- and θ-spending sequences.

Remark: Tsiatis and Mehta (2003) argue that every adaptive trial can be formulated as a group sequential test $(\mathscr{R}_j, \mathscr{A}_j)$, $j = 1, \ldots, K$, with some specific α and θ-sequence and hence is uniformly improvable by the corresponding sequential likelihood ratio test. However, in order to use the sequential likelihood ratio test the sample size adaptation rule and the α- and θ-spending sequence have to be known a priori and cannot be adapted in the course of the trial. Similarly, the power is prefixed in such a group sequential design. However, the choice of the sample size functions, α-, θ-spending and power functions is a question of the costs for false positive and false negative decisions (and the gain of correct decisions), as well as the costs of the interim analyses and the sample sizes: minimizing specific costs will lead to specific α- and θ-spending and power functions. The costs of a clinical trial depend on economic and medical factors which are usually difficult to quantify even more at the beginning of the trial. The costs might become clearer in the course of the trial as knowledge on these factors increase. Formally, adaptive designs allow one to modify sample sizes and the α-, θ-spending and power functions accordingly.

3.2. The resulting flexible design

We now show how the sequential likelihood ratio test can be modified mid-trial using all the information collected so far without compromising the prespecified overall level α. Assume that an unscheduled adaptation is done at stage $j < K$ based on the information from stages 1 to j. We denote $l_1, \ldots, l_j$ the observed values of the likelihoods $L_1, \ldots, L_j$ and assume that $a_i \leqslant l_i \leqslant b_i$ for $i = 1, \ldots, j$, i.e., no stopping condition has been raised until stage j. The conditional type I error probability of the sequential likelihood ratio test at stage j is $A_j^{\mathrm{LR}}(l_1, \ldots, l_j) = P_0(\bar{\mathscr{R}}_K^{\mathrm{LR}} \mid L_1 = l_1, \ldots, L_j = l_j)$. Since $a_t \leqslant l_t \leqslant b_t$ for all $t \leqslant j$ this is equal to $A_j^{\mathrm{LR}}(l_1, \ldots, l_j) = P_0(\bar{\mathscr{R}}_{K|j}^{\mathrm{LR}} \mid L_1 = l_1, \ldots, L_j = l_j)$ with $\bar{\mathscr{R}}_{K|j}^{\mathrm{LR}} = \cup_{i=j+1}^{K} \mathscr{R}_i^{\mathrm{LR}}$. The follow-

ing notation will be useful in the sequel. For $i > j$ let $\bar{L}_{i|j} = \Pi_{s=j+1}^{i} L_s$, $b_{i|j} = b_i / \Pi_{t=1}^{j} l_t$ and $a_{i|j} = a_i / \Pi_{t=1}^{j} l_t$. Notice that $b_{i|j}$ and $a_{i|j}$ depend on the data observed until stage j via the likelihood $\bar{l}_j = \Pi_{t=1}^{j} l_t$. Since $a_t \leqslant l_t \leqslant b_t$ for all $t \leqslant j$ if follows that $\mathscr{R}_{j+1}^{\mathrm{LR}} = \{L_{j+1} > b_{j+1|j}\}$ and $\mathscr{R}_i^{\mathrm{LR}} = \{a_{j+1|j} \leqslant \bar{L}_{j+1|j} \leqslant b_{j+1|j}, \ldots, a_{i-1|j} \leqslant \bar{L}_{i-1|j} \leqslant b_{i-1|j}, \bar{L}_{i|j} > b_{i|j}\}$ for $i > j+1$. Hence $A_j^{\mathrm{LR}}(l_1, \ldots, l_j)$ is the type I error probability of a sequential likelihood ratio test which starts at stage $j+1$, has at most $K-j$ stages, and rejection and acceptance boundaries $a_{i|j}$ and $b_{i|j}$, $i = j+1, \ldots, K$.

According to (1) the trial can be continued with any test with conditional level $A_j(l_1, \ldots, l_j)$. Of course, the originally planned test satisfies this condition.

3.3. Conditionally optimal flexible design

As an example for an adaptation, consider the case where the preplanned α- and θ-spending sequences are altered at the interim analysis. To this end we introduce a conditional version of the spending sequences. For $i = j+1, \ldots, K$ let $\alpha_{i|j} = \alpha_{i|j}(l_1, \ldots, l_j) = P_0(\bar{\mathscr{R}}_i^{\mathrm{LR}} \mid l_1, \ldots, l_j)$ and $\theta_{i|j} = \theta_{i|j}(l_1, \ldots, l_j) = P_0(\bar{\mathscr{A}}_i^{\mathrm{LR}} \mid l_1, \ldots, l_j)$ denote the sequences of the conditional rejection and acceptance probabilities for the remainder of the preplanned sequential likelihood ratio test. We refer to them as the conditional α- and θ-spending sequence. Obviously, $\theta_{i|j} + \alpha_{i|j} < 1$ for $i = j+1, \ldots, K-1$ and $\theta_{K|j} + \alpha_{K|j} = 1$. Further, there is a simple relationship between the conditional and unconditional α- and θ-sequences, namely, for all $i > j$, $\alpha_i = P_0(\cup_{t=1}^{j} \mathscr{R}_t) + E_0[\alpha_{i|j}(L_1, \ldots, L_j)]$ and $\theta_i = P_0(\cup_{t=1}^{j} \mathscr{A}_t) + E_0[\theta_{i|j}(L_1, \ldots, L_j)]$, where expectation is taken over $L_1, \ldots, L_j$.

The new sequential design for the remainder of the trial can be fixed based on the unblinded interim data. It might have a larger or smaller maximum number of inspection times $\tilde{K}$, altered sample sizes per stage, and some new conditional α- and θ-spending sequences, say $\tilde{\alpha}_{j+1|j} \leqslant \cdots \leqslant \tilde{\alpha}_{\tilde{K}|j}$, and $\tilde{\theta}_{j+1|j} \leqslant \cdots \leqslant \tilde{\theta}_{\tilde{K}|j}$ with $\tilde{\theta}_{i|j} + \tilde{\alpha}_{i|j} < 1$ for $i < \tilde{K} - 1$ and $\tilde{\theta}_{\tilde{K}|j} + \tilde{\alpha}_{\tilde{K}|j} = 1$. According to (1) the only condition required for a control of the overall type I error rate is $\tilde{\alpha}_{\tilde{K}} \leqslant A_j^{\mathrm{LR}}(l_1, \ldots, l_j)$. In general, a change of the conditional spending sequences will cause also a change of the unconditional spending sequences.

Given the new conditional α- and θ-sequences $\tilde{\alpha}_{i|j}$ and $\tilde{\theta}_{i|j}$ $(i > j)$, we can still use the likelihood ratios of the forthcoming stages by applying appropriate (conditional) acceptance and rejection boundaries $\tilde{a}_{i|j}$ and $\tilde{b}_{i|j}$ for $i > j$. The new conditional boundaries can be determined as the boundaries of a sequential likelihood ratio test (cf. the end of Section 3.1) but now using conditional rejection probabilities instead of unconditional rejection probabilities. According to Theorem 1 of Tsiatis and Mehta (2003), the resulting test is optimal conditional on the interim data and the adaptively chosen conditional α- and θ-sequence.

4. Concluding remarks

In general, adaptive tests are no sequential likelihood ratio tests. This is the case also for the adaptive test described above, although it is built up using the likelihood ratios

of the individual stages. According to Theorem 1 in Tsiatis and Mehta (2003), choosing the "right" spending function from the beginning would allow for a more efficient test design. However, such a choice may not be possible in practice. Moreover, for a design with sample size recalculation, the suggestion of Tsiatis and Mehta (2003) means to replace an adaptive design by a sequential design which has a substantially larger number of interim analyses. This is acceptable only in cases where the costs of an interim analysis are small. Interim analyses require statistical and data management resources. They can lead to a substantial delay in time or face the problem of overrunning patients which cannot be included in the interim analysis. Unblinding the data mid-trial also bears the danger of corrupting the integrity of the trial because of a leakage of interim results to investigators and others involved in the trial. For these reasons many clinical trials have a rather small maximum number of interim analyses. Posch et al. (2003) demonstrate that adaptively inserting or skipping interim analyses in an adaptive design can provide an efficient strategy to save interim analyses without increasing the average sample size. Another important issue for adaptive designs is the control of conditional power which requires sample size recalculations. Brannath and Bauer (2004) give examples from applications where a control of conditional power is essential, and they identify two stage designs which minimize the expected sample size while keeping the overall level, the overall power and the conditional power at prespecified levels. It turns out that such designs differ from group sequential designs and from adaptive designs which use the conditional error function of a group sequential design.

References

Bauer, P., 1989. Multistage testing with adaptive designs (with discussion). Biometrie Inform. Med. Biol. 20, 130–148.

Bauer, P., Köhne, K., 1994. Evaluation of experiments with adaptive interim analyses. Biometrics 50, 1029–1041.

Bauer, P., Posch M., 2004. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections by H. Schäfer and H.-H. Müller, Statistics in Medicine 2001; 20: 3741–3751. Statist. Med. 23, 1333–1334.

Brannath, W., Bauer, P., 2004. Optimal conditional error functions for the control of conditional power. Biometrics 60, 715–723.

Lan, K.K.G., DeMets, D.L., 1983. Discrete sequential boundaries for clinical trials. Biometrika 70, 659–663.

Liu, Q., Proschan, M.A., Pledger, G.W., 2002. A unified theory of two-stage adaptive designs. J. Amer. Statist. Assoc. 97, 1034–1041.

Müller, H.-H., 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.

Müller, H.-H., Schäfer, H., 2004. A general statistical principle for changing a design any time during the course of a trial. Statist. Med. 23, 2497–2508.

Posch, M., Bauer, P., 1999. Adaptive two stage designs and the conditional error function. Biometrical J. 41, 689–696.

Posch, M., Bauer, P., Brannath, W., 2003. Issues on flexible designs. Statist. Med. 22, 953–969.

Proschan, M.A., Hunsberger, S.A., 1995. Designed extension of studies based on conditional power. Biometrics 51, 1315–1324.

Schäfer, H., Müller, H.-H., 2001. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. Statist. Med. 20, 3741–3751.

Tsiatis, A.A., Mehta, C., 2003. On the inefficiency of the adaptive design for monitoring clinical trials. Biometrika 90, 367–378.