A New Class of Self-Designing Clinical Trials

Joachim HARTUNG

Department of Statistics[*]*, University of Dortmund,*

D-44221 Dortmund, Germany

Address:

Department of Statistics

University of Dortmund

D-44221 Dortmund, Germany

Phone: ++49 231 755 3163

Fax: ++49 231 755 5304

Email: Hartung@statistik.uni-dortmund.de

---

[*] and project 'Meta-Analysis in Biometry and Epidemiology' (SFB 475) of the German Research Community (DFG)

A New Class of Self-Designing Clinical Trials

Abstract. A class of self-designing clinical trials is considered which according to an effective but simple, finite learning algorithm consists of automatically adaptively planned weighted group sequential trials with a decision about rejection of the null-hypothesis at each step, but the full level-α-test at the end of the study preserved.

1. Introduction

In a recent paper [1] self-designing clinical trials are introduced in a general setting for normal variables with known variances, for which in [2] a concrete proposal for building the test statistic is given. There the sequence of sample sizes is fixed prior to the beginning of the study. In [3] based on a general learning scheme completety self-designing trials are considered, using the inverse normal method, cf. [4], for transforming the p-values associated with the respective test statistics in the sequential groups.

Here now in defining our easily to handle learning rule for a finite self-designing we employ the whole prominent family of χ^2-distributions to derive our transformations of the p-values, having desirable properties in our context, cf. [5]. Thus we get a new class of weighted self-designing trials allowing for sequential decisions about the null-hypothesis, contrarily to the class discussed above, but also with the full level-α-test at the end of the trial preserved.

With respect to an early use of p-values and adaptive sample size planning in this framework let us refer to [6], [7], [8], [9].

2. Basic statistics

For some real valued parameter ϑ we consider the problem of testing at prescribed size α the hypothesis $\mathrm{H}_0 : \vartheta = 0$ vs. $\mathrm{H}_1 : \vartheta > 0$, where for example $\vartheta = \theta_1 - \theta_2$ if θ_1, θ_2 denote the expectations of the outcome variables of e.g. 'verum' and 'placebo', respectively, in a controlled clinical study.

The study is devided into several, disjoint study parts: $stp(k)$, $k = 1,\ldots,K$. Upon the decision rule not all study parts are necessarily carried out.

In $stp(k)$ let T_k be a one-sided test statistic for testing H_0 vs. H_1, where large values of T_k lead to a rejection of H_0. Under H_0 may T_k have a continuous distribution function $F_{k,0}$.

Then under H_0 the p-value $p_k = 1 - F_{k,0}(T_k)$, $k = 1,\ldots,K$, is uniformly distributed on the interval (0, 1), and $q_k(v_k) = F^{-1}_{\chi^2(v_k)}(1 - p_k)$, $k = 1,\ldots,K$, belongs to a (central) χ^2-distribution with v_k degrees of freedom, where $F^{-1}_{\chi^2(v_k)}$ denotes the inverse of the $\chi^2(v_k)$-distribution function, i.e. $q_k(v_k)$ is the $(1 - p_k)$-quantile $\chi^2(v_k)_{1-p_k}$ of the $\chi^2(v_k)$-distribution, cf. [4], [5].

For the combined statistic up to $stp(k)$ thus we get under H_0

$$S_k = \sum_{j=1}^{k} q_j(v_j) \overset{\mathrm{H}_0}{\sim} \chi^2(v_\Sigma(k)), \quad v_\Sigma(k) = \sum_{j=1}^{k} v_j, \quad k = 1,\ldots,K.$$

Hence in particular, if

$$S_k \geq \chi^2(v_\Sigma(K))_{1-\alpha}, \quad \text{for some } k \in \{1,\ldots,K\},$$

so H_o is rejected at size α, since $S_K \geq S_k$, because of $q_j(v_j) \geq 0$.

If T_k has under H_0 an exact distribution that is only 'nearly continuous' or if its distribution is approximated by a continuous distribution function, then the above results hold in a corresponding good approximation.

Note that for $v_j = 2$ we get by S_k R.A. Fisher's combining method, cf. [4], [5], [6], [7].

3. Self-designing

At the beginning of the study we decide for a (fictitious) maximum number K of study parts, *stp*(*k*), respectively number *K*-1 of interim analyses, a minimum number $v_{\min}$ of degrees of freedom to be used in a realized *stp*(*k*), say $v_{\min} = 1$ here, and we put $v_{\Sigma}(K) = K$, implying the global critical value: $\mathrm{cv}_{\alpha} = \chi^2(K)_{1-\alpha}$.

For a chosen type II error rate β there may exist a sample size spending function $f = f(T)$ such that for the use of the test statistic T by: $n = f_{k-1}(\alpha,\beta)$, the minimum realistic sample size n for a trial is delivered holding the type I and II error rates α and β conditionally under the knowledge available up to $stp(k-1)$, where $stp(0)$ stands for the apriori information.

Further we have to choose a starting configuration of sample size and degrees of freedom, n_1, v_1 in $stp(1)$, and a real valued relaxation parameter κ; influencing the number of study parts to be carried out really.

Let now up to *stp*(*k-1*) the test values T_j , respectively p_j and $q_j(v_j)$, be given, and if $S_{k-1} \geq \mathrm{cv}_{\alpha}$, then H_0 is rejected at level α, and the trial stops.

Otherwise the sample size n_k and the associated degrees of freedom v_k for *stp*(*k*) are determined as follows. Let be

$$M_k = f_{k-1}(1 - F_{\chi^2(K - v_{\Sigma}(k-1))}(\mathrm{cv}_{\alpha} - S_{k-1}),\ \beta),$$

i.e. M_k would be that sample size needed for holding α and β in just one further *stp* by associating the full remaining degrees of freedom, conditionally under the results obtained till that time.

But since the parameter estimates involved in the planning function f_{k-1} may not yet have stabilized, only a part of M_k should be in general the size of the next *stp*. We give now a learning rule that represents an effective but simple way of self-designing.

Define

$$v_k^* = \frac{v_\Sigma(k-1)+\kappa}{K+\kappa}\cdot\{K - v_\Sigma(k-1)\}\ ,$$

then the degrees of freedom to be associated with *stp*(*k*) let be given by

$$v_k = \begin{cases} [v_k^* + 1], & \text{if } \mathrm{K} - v_k^* - v_\Sigma(k-1) \geq 1 \\ K - v_\Sigma(k-1), & \text{otherwise} \end{cases},$$

where $[v_k^* + 1]$ denotes here the largest natural number less than $v_k^* + 1$. Hence the sample size of *stp*(*k*) may be determined as

$$n_k = \frac{v_k}{K - v_\Sigma(k-1)}\cdot M_k\ ,$$

respectively as the smallest realistic size greater than this value.

The trial stops after *stp*(k^*) if:

$$S_{k^*} = \sum_{j=1}^{k^*} q_j(v_j) \geq \mathrm{cv}_\alpha,\ \text{ or if: } n_{k^*} = M_{k^*}\ ;\ (\text{putting } n_i = v_i = 0 \text{ if } \mathrm{k}^* < \mathrm{i} \leq \mathrm{K}).$$

If $S_{k^*} \geq \mathrm{cv}_\alpha$, we reject H_0 at level α, otherwise we stay with H_0.

If the sequence of the M_k is not markedly decreasing after some steps the trial may be stopped. We can also introduce a lower bound for an early acceptance of H_0, i.e. if for example $S_k \leq \chi^2(v_\Sigma(k))_{\alpha_L}$, where α_L should not be chosen too conservative in order not to cut paths early that would lead to a rejection of H_0, cf. also [2]. Furthermore, the updating parameter κ can be chosen in dependence on *'k-1'*.

The adaptive planning of *stp*(*k*) by using information from the previous study parts does not affect the independence of $q_k(v_k)$ and S_{k-1} under H_0, $q_k(v_k)$ is in any case $\chi^2(\nu_k)$-distributed under H_0; cf. the quite analogous argumentations in [1], [6], [7], [8], [9].

4. Example

Let us consider two medications with binary outcomes, and θ_1, θ_2 be for instance the expected cure rates. We are interested in the one-sided test problem: H_0: $\theta_1 = \theta_2$ vs. H_1: $\theta_1 > \theta_2$.

For the test statistic T in the (2×2)-table analyses of the various study parts we take the (one-sided) χ^2-test, with the p-value $p = (1 - F_{\chi^2(1)}(T))/2$, as long as for the usual estimates $\hat{\theta}_1 > \hat{\theta}_2$ holds, otherwise the one-sided underlying normal test or, of course, in any case the Fisher-Irwin test can be taken, and sample size calculations for determining M_k can be done by use of tables or approximate formulas, e.g. [10], [11, p. 418-421], which are also to find in software packages.

To get a short sequence of study parts the parameters in our example are chosen as follows: $\alpha = 0.025$, $\beta = 0.10$, $K = 10$, $\kappa = (k\text{-}1) \cdot 4$, and having no real prior information, but only a guess that the total number of patients to be involved in the study will lie between 150 and 250, we take as starting configuration: $n_1 = 40$ and $v_1 = 2$ in *stp*(1).

So the weights of the following study parts in form of degrees of freedom are: $v_2 = 4$, $v_3 = 4$; i.e. the trial consists really of at most three parts. The global critical value is given by: $cv_{0.025} = \chi^2(10)_{0.975} = 20.5$.

Now the trial starts, and we observe in *stp*(1) the test value of T as: $T_1 = 1.67$, being the 0.8- = (1-2·p_1)-quantile $\chi^2(1)_{0.8}$ of $\chi^2(1)$, thus: $p_1 = 0.1$, and $q_1(2) = \chi^2(2)_{0.9} = 4.6 = S_1$.

Further we extract from *stp*(1): $\hat{\theta}_1 = 0.7$ and $\hat{\theta}_2 = 0.5$, such that by: $cv_\alpha - S_1 = 15.9 \approx \chi^2(8)_{0.95} = F^{-1}_{\chi^2(8)}(0.95)$, a sample size planning for a one-sided (2×2)-table analysis with a type I error rate 0.05 and $\beta = 0.1$ under assuming the above cure rate estimates yields: $M_2 = 222$, or $n_2 = 112$, (111).

If these estimates for θ_1, θ_2 stay constant in *stp*(2), we observe $T_2 = 4.7$ implying: $p_2 = 0.015$, $q_2(4) = \chi^2(4)_{0.985} = 12.5$, $S_2 = 17.1$, and $cv_\alpha - S_2 = 3.4 = \chi^2(4)_{0.50} = F^{-1}_{\chi^2(4)}(0.50)$. Hence sample size planning as above, however with a type I error rate of 0.50 now, yields: $n_3 = M_3 = 46$.

With the same treatment effects we observe $T_3 = 1.9$ yielding: $p_3 = 0.08$, $q_3(4) = \chi^2(4)_{0.92} = 8.5$, and so: $S_3 = 25.6 > cv_{0.025}$, i.e. H_0 is rejected at level $\alpha = 0.025$.

Note that if with a larger treatment difference we had got $S_2 > 20.5$, then the trial had stopped already after *stp*(2) with a rejection of H_0 at the same level.

We needed about 200 patients to be enclosed in the whole study, assuming an ideal situation here for demonstrational purpose. In a fixed sample size plan we would have calculated about 270 necessary patients, if the estimates for the cure rates would have been known in advance.

Hence, if treatment effects remain nearly stable during the sequence of trials, - otherwise all designing procedures can come into troubles -, so we can say that our learning method in designing the trial adaptively uses patients sparingly.

5. Multi-centre trials

Since after *stp*(*k*-1) the remaining total number of degrees of freedom is fixed we may also choose $v_k \in \{1,2,\ldots,K - v_\Sigma(k-1)\}$ under using information from the previous (,- which has to be assured, of course, e.g. by a fixed rule or by an independent data-monitoring committee -,) study parts without affecting under H_0 the independence of q_k and S_{k-1}, or the distribution of q_k, cf. also e.g. [1]. Now in multi-centre trials, for instance, such a case may occur as follows.

Because of possible centre effects, concomitant variables and organization effort with an interim analysis, a minimum number $n_{\min}$, possibly in dependence on *k*, for the sample size of *stp*(*k*) seems to be useful. So with n_k and v_k from sec. 3 we may define the modified quantities by

$$n_{k,\mathrm{mod}} = \begin{cases} n_k, & \text{if } n_k \geq n_{\min} \\ n_{\min}, & \text{otherwise} \end{cases}, \quad \text{and} \quad v_{k,\mathrm{mod}} = (K - v_\Sigma(k-1)) \cdot \frac{n_{k,\mathrm{mod}}}{M_k} \ ,$$

respectively the next natural number not greater than $K - v_\Sigma(k-1)$.

Similarly to [2] and [3] also more complicated self-designing rules can be formulated. With a suitable reformulation, cf. [3], two-sided hypotheses regarding several treatment arms can be included, too, taking the correct interpretation of the underlying composite hypothesis into consideration.

6. Concluding remark

We propose a flexible and effective learning method that allows for a completely self-designing of a group sequential trial, with a decision about rejection of H_0 at each step. Due to its ease of construction it is simple to apply.

The termination of the study is steered by weighting the sequence of sub-trials in form of choosing different χ^2-distributions for transforming the p-values.

Since at the end of the trial a full level-α-test is preserved, the proposed method is appealing for investigators, who usually are difficult to convince to pay a price for an interim look, cf. also [6], [7].

References

[1] L. Fisher, Self-designing Clinical Trials, *Statistics in Medicine* **17** (1998) 1551-1562.

[2] Y. Shen and L. Fisher, Statistical Inference for Self-designing Trials with a One-sided Hypothesis, *Biometrics* **55** (1999) 190-197.

[3] J. Hartung, A Self-designing Rule for Clinical Trials with Arbitrary Response Variables, *Preprint*, SFB 475, University of Dortmund.

[4] L.V. Hedges and I. Olkin, *Statistical Methods for Meta-Analysis*, Academic Press, Orlando, 1985.

[5] J.I. Marden, Sensitive and Sturdy p-values, *The Annals of Statistics* **19** (1991) 918-934.

[6] P. Bauer and K. Köhne, Evaluation of Experiments with Adaptive Interim Analyses, *Biometrics* **50** (1994) 1029-1041.

[7] P. Bauer and J. Röhmel, An Adaptive Method for Establishing a Dose-response Relationship, *Statistics in Medicine* **14** (1995) 1595-1607.

[8] M.A. Proschan and S.A. Hundsberger, Designing Extension of Studies Based on Conditional Power, *Biometrics* **51** (1995) 1315-1324.

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

[10] J.K. Hasemann, Exact Sample Sizes for Use with the Fisher-Irwin Test for (2×2)-tables, *Biometrics* **34** (1978) 106-109.

[11] J. Hartung, B. Elpelt and K.H. Klösener, *Statistik: Lehr- und Handbuch der angewandten Statistik*, Oldenbourg Verlag, München, Wien, 12th ed., 1999.