Adaptive Treatment Selection with Survival Endpoints

Gernot Wassmer

Institut für Medizinische Statistik, Informatik und Epidemiologie
Universität zu Köln

Joint work with Markus Roters, Omnicare Clinical Research, Cologne

BBS, EFSPI
Seminar on Adaptive Designs in Drug Development
Basel 15th June, 2007

„Adaptive“ means:

Planning of subsequent stages may be based on information observed so far, under control of an overall Type I error rate (at least approximately).

Particularly in survival designs:

- Reassessment of necessary number of events
- Selecting treatment arms
- Adaptive choice of test statistic
- Changing the number of interim analyses, type of stopping boundaries, etc.

Recent literature on adaptive design in survival trials:

- Schäfer and Müller (Statistics in Medicine, 2001)
- Lawrence (J. Biopharm. Stat., 2002)
- Shen and Cai (J. Amer. Stat. Ass., 2003)
- Li, Shih, and Wang (ASA Proceedings, 2003)
- Bauer and Posch (Statistics in Medicine, 2004)
- Wassmer (Biometrical Journal, 2006)

Adaptive design using the *inverse normal method*

Consider at *k*th stage

$$T_k^* = \left(\sum_{\tilde{k}=1}^{k} w_{\tilde{k}}^2\right)^{-1/2} \sum_{\tilde{k}=1}^{k} w_{\tilde{k}}\, \Phi^{-1}(1-p_{\tilde{k}}),$$

where $w_1, w_2, \ldots, w_K$ are weights fixed prior to the trial.

Specifically:

Stagewise, under H_0 standard normally distributed variables

$$Z_k = \Phi^{-1}(1-p_k)$$

are necessary for using this method.

Test for Means or Test for Rates

Fix weights through

$$w_k = \sqrt{n_k}\ , k = 1,...,K,$$

where $n_1,\ldots,n_K$: planned sample sizes.

With these weights the adaptive test coincides with the classical group sequential test if no adaptations were performed.

Survival Design

Consider two treatment groups testing

$$H_0 : \omega = \frac{\ln(1-\pi_2)}{\ln(1-\pi_1)} = 1,$$

where ω is the hazard ratio and π_1 and π_2 denote the event rates in the two treatment groups.

Considering non-inferiority or equivalence testing is also possible, i.e.,

$H_0 : \omega \leq \delta_0$ (noninferiority)

$H_0 : \omega \leq \delta_0$ or $\omega \geq \frac{1}{\delta_0}$ (equivalence)

Survival Design

During the stages of the trial, a sequence of accumulated events $d_1,\ldots,d_K$ is observed.

At each stage k of the test procedure the logrank test statistic

$$LR_k^* = \frac{\sum_{i=1}^{d_k}\left(I_{2i} - \frac{N_{2ik}}{N_{1ik}+N_{2ik}}\right)}{\sqrt{\sum_{i=1}^{d_k}\frac{N_{1ik}N_{2ik}}{(N_{1ik}+N_{2ik})^2}}}, k = 1,\ldots,K,$$

is calculated, where $I_{2i} = 1$ if the ith event occurs in treatment 2, and N_{1ik} and N_{2ik} are the number of patients at risk at stage k in treatment groups 1 and 2, respectively, when the ith event occurred.

Survival Design

Approximately, for fixed d_k, LR_k^* has unit variance and

$$E(LR_k^*) = \sqrt{d_k}\,\frac{\sqrt{r}}{1+r}\log(\omega)\,, k = 1,...,K,$$

where ω is the unknown hazard ratio and $r = N_2/N_1$ is the allocation rate.

Approximately, the sequence of test statistics $LR_1^*,...,LR_K^*$ has the independent and normally distributed increments structure.
(Jones and Whitehead, 1979; Sellke and Siegmund, 1982; Tsiatis, 1981, 1982)

Therefore, the group sequential tests can be applied in the usual way.

Survival Design

Test statistic of adaptive test

Adaptations possible when using the combination test statistic

$$T_k^* = \left(\sum_{\tilde{k}=1}^{k} w_{\tilde{k}}^2\right)^{-1/2} \sum_{\tilde{k}=1}^{k} w_{\tilde{k}} Z_{\tilde{k}} \text{ , where}$$

$$Z_k = \frac{\sqrt{d_k} LR_k^* - \sqrt{d_{k-1}} LR_{k-1}^*}{\sqrt{d_k - d_{k-1}}} :$$ approximately independent and normally distributed increments

Survival Design

Fix weights through

$$w_1 = \sqrt{\zeta_1} \quad \text{and} \quad w_k = \sqrt{\zeta_k - \zeta_{k-1}}\,, k = 2,...,K,$$

where ζ_k: planned (or expected) number of accumulated events at stage *k.*

Note

If $d_k = \zeta_k$, then $T_k^* = LR_k^*$

Planning Tools

As for group sequential designs, i.e.

1. Calculate number of events d_f in a fixed sample size design

2. $D = d_K$ in the group sequential design is approximated by

   $$D = I(K, \alpha, \beta)\, d_f \ ,$$

   where I is the inflation factor referring to a specific group sequential test design.

Planning Tools

3. Estimate number of patients through

   $$N = D / \Psi(a+f) ,$$

   where $\Psi(s)$: Probability of an event at time s.

4. Estimate the observation times $s_1,\ldots,s_K$ such that the expected information rates are proportional to a sequence of specified information levels $\zeta_1,\ldots,\zeta_K$ (Kim & Tsiatis, 1990).

Consider *G* treatment arms and one control

Example *G* = 3, equal sample sizes between the treatment groups

$$H_0^1 : \omega_1 = \frac{\ln(1-\pi_1)}{\ln(1-\pi_0)} = 1,$$

$$H_0^2 : \omega_2 = \frac{\ln(1-\pi_2)}{\ln(1-\pi_0)} = 1,$$

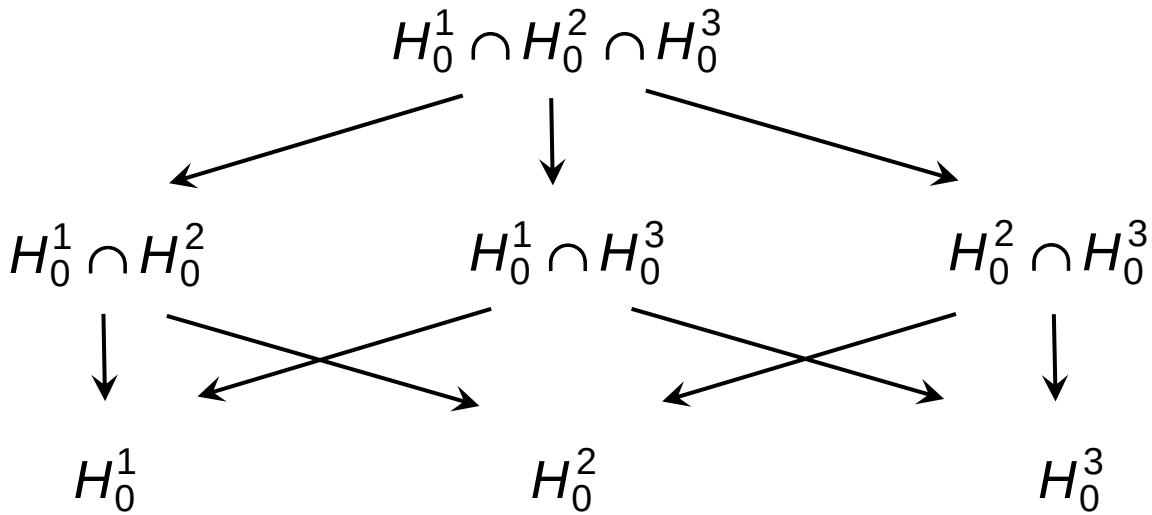
$$H_0^3 : \omega_3 = \frac{\ln(1-\pi_3)}{\ln(1-\pi_0)} = 1.$$

Assume that one treatment arm is to be selected at the first interim stage. Confirmatory analysis should be possible for the comparison of the selected dose with the control, H_0^S.

Adaptive Treatment Selection

Adaptive treatment selection is based on the application of the closed test procedure together with combination tests (e.g., Bauer & Kieser, 1999; Posch et al., 2005). This guarantess strong control of the Type I error rate.

Closed system of hypotheses:



Test Procedure

At the first interim analysis, consider the test statistic

$$LR_1^* = \max_{i=1,2,3}(LR_1^{1*}, LR_1^{2*}, LR_1^{3*}),$$

where LR_1^{i*} denotes the logrank statistic for H_0^i, $i = 1,2,3$.

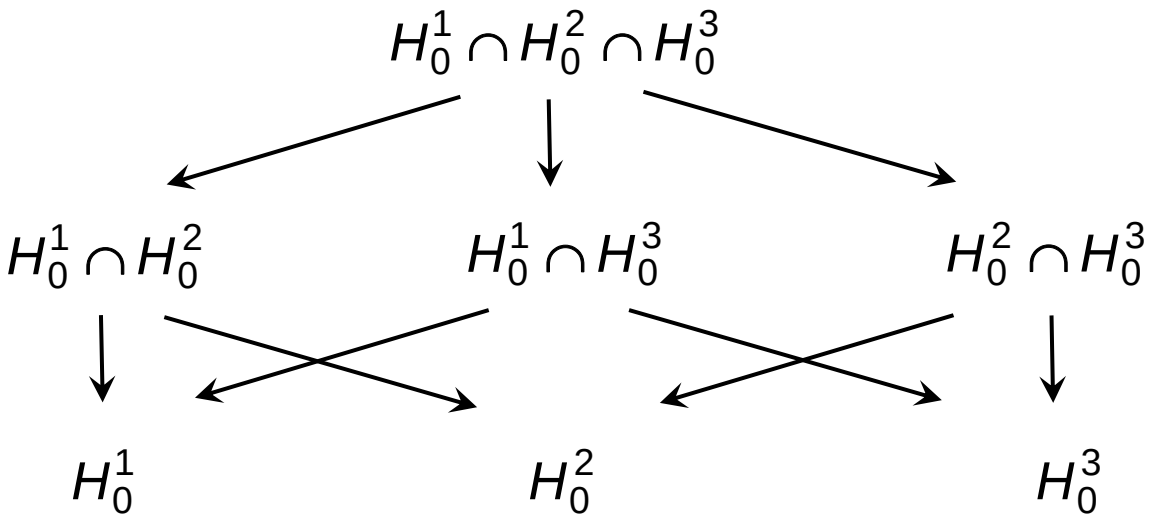
This test statistic has, asymptotically, the distribution of the maximum of vector of correlated standard normally distributed random variables (with common correlation coefficient 0.5 due to equal group sizes among the treatment arms).
Thus, Dunnett's critical values can be used.

See also Follmann, Proschan, and Geller (1994), who derive the asymptotic normality of the joint distribution of all log-rank statistics (evaluated at fixed study times) in a pairwise comparison setting.

Test Procedure

- At the first interim analysis, it is possible to stop the trial while showing significance of one (or more) treatment arms.
- It is also possible to stop the trial due to futility arguments. These are usually based on conditional power calculations.
- It is expected, however, that the first stage is specifically used to select a treatment arm to be considered in the subsequent stages of the trial and to reassess the sample size (number of events) for the subsequent stages.

Stage I **Stage II** **…**



?

?

H_0^S

Simple “trick”: Test of intersection hypotheses are formally performed as tests for H_0^S.

Let $\varphi(p,q) = \dfrac{w_1\Phi^{-1}(1-p) + w_2\Phi^{-1}(1-q)}{\sqrt{w_1^2 + w_2^2}}$

Test decision for the second stage:

H_0^S is rejected if

$$\min_{J \ni S} \varphi(p_J, q_S) \geq u_2 ,$$

where p_J is the p-value of the Dunnett test for testing $\bigcap_{i \in J} H_0^i$,

q_S is the independent increment of the test statistic for the selected dose, i.e.,

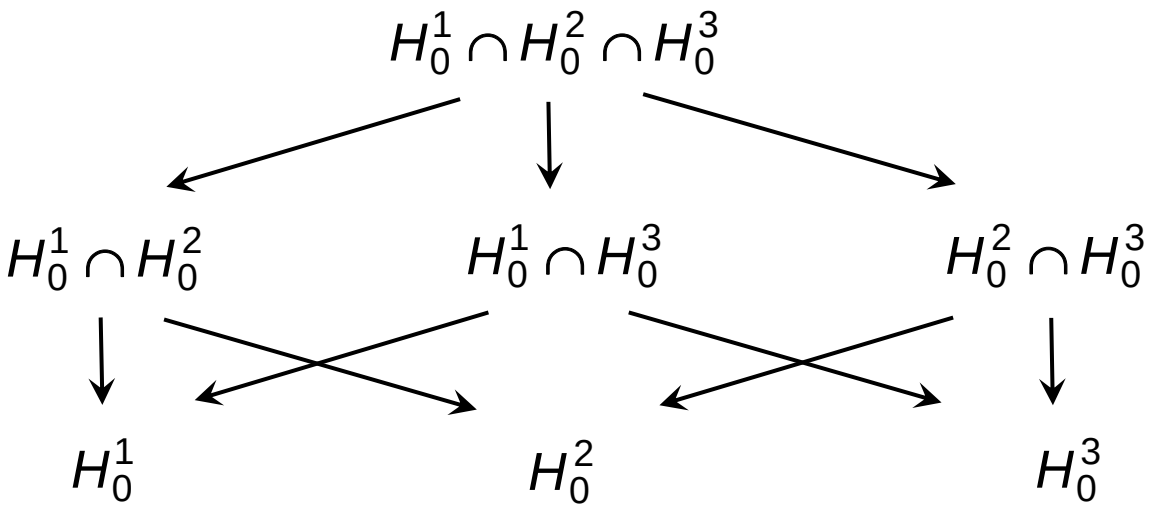
$$q_S = 1 - \Phi\left(\frac{\sqrt{d_2^S} LR_2^{S*} - \sqrt{d_1^S} LR_1^{S*}}{\sqrt{d_2^S - d_1^S}}\right),$$

and u_2 is the critical value for the second stage.

Example $S = 3$

Stage I

Stage II **…**

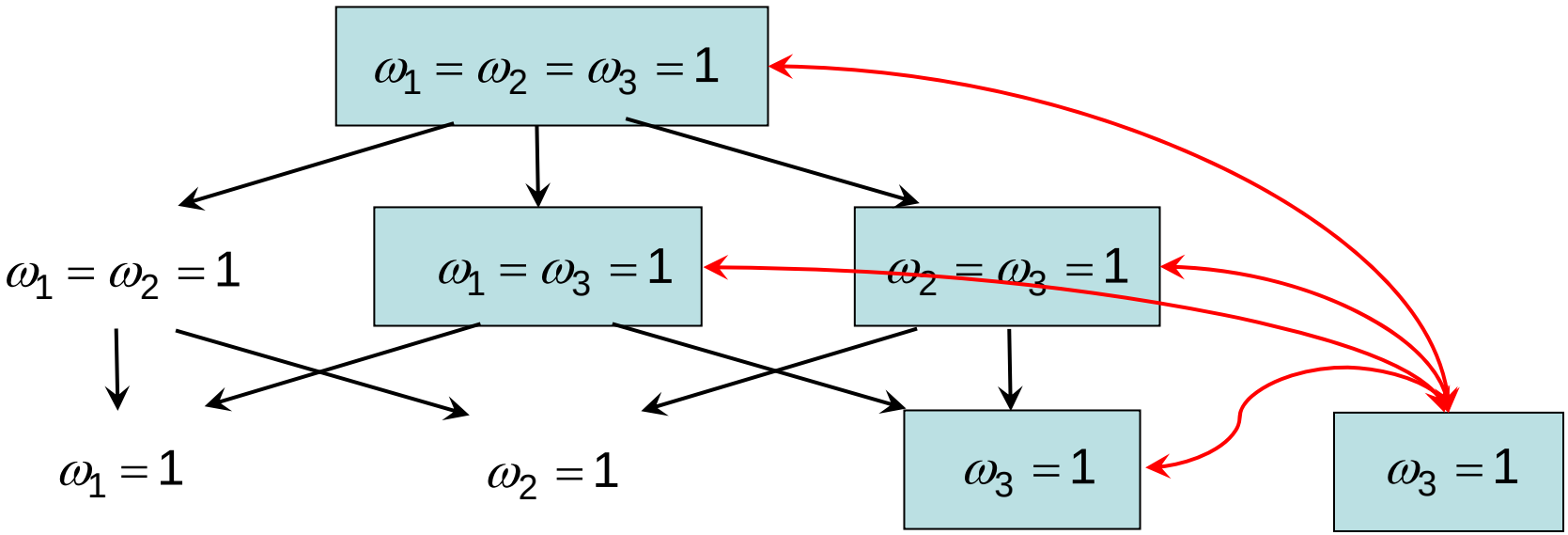


H_0^3

Note:
If the treatment arm with the largest test statistic is selected, it suffices to combine the test for H_0: $\omega_1 = \omega_2 = \omega_3 = 1$ with the test for H_0: $\omega_3 = 1$

Example *S* = 3

Stage I **Stage II** **…**



H_0^3 can be rejected if all combination tests exceed the critical value u_2 .

Discussion

For control of type I error rate, it is generally necessary that

- Rules for early rejection of hypotheses are stated in the protocol
- Rules for combining evidence across both stages and associated multiplicity adjustments are defined up front (in the protocol)
- For regulatory purposes, the class of envisaged decisions after stage 1 is stated (in the protocol)
- The “rules” for adaptation and stopping for futility
  - not need to be pre-specified
  - can make use of Bayesian principles integrating all information available, also external to the study
  - should be evaluated (e.g. via simulations) and preferred version recommended, e.g., in DMC charter

Discussion

- Treatment selection can also be based on surrogate parameter, i.e., switch of endpoint can be reasonable.
- You may also use partial ML estimate for treatment difference from Cox model to adjust for cofactors (Jennison and Turnbull, 2000)
- Test procedure can be generalized to selection of more than one treatment arm
- It is also possible to add new treatment arms

Discussion

- Alternatives (König et al., 2007, Posch et al., 2005)
  - Different tests for intersection hypotheses (Bonferroni, Sidak, Simes)
  - Classical Dunnett test
  - Conditional error Dunnett test
- Software should be available for assessment of the procedures through simulation
- Estimation procedure should be available

References

- Bauer, P., Kieser, M. (1999). Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine* 18, 1833-1848.
- Bauer, P., Posch, M. (2004). Letter to the Editor: Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Statistics in Medicine* 23: 1333-1335.
- Bretz, F., Schmidli, H., König, F., Racine, A., Maurer W. (2006) Confirmatory seamless Phase II/III clinical trials with hypothesis selection at interim: General concepts. *Biometrical Journal* 2006; 48, 623-634
- Follmann, D.A., Proschan M.A., Geller, N.L. (1994). Monitoring pairwise comparisons in multi-armed clinical trials. *Biometrics* 50, 325-336.
- Jones, D., Whitehead, J. (1979). Sequential forms of the logrank and modified Wilcoxon tests for censored data. *Biometrika* 66: 105–133.
- Kim, K., Tsiatis, A. A. (1990). Study duration for clinical trials with survival response and early stopping rule. *Biometrics* 46: 81-92.
- König, F., Brannath, W., Bretz, F., Posch, M. (2007) Adaptive Dunnett tests for treatment selection. Submitted.
- Lawrence, J. (2002). Strategies for changing the test statistic during a clinical trial. *J. Biopharm Statist.* 12, 193-205.
- Lehmacher, W., Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* 55, 1286-1290.
- Li, G., Shih, W. J., Wang, Y. (2003). Two-stage adaptive designs for clinical trials with survival data. ASA Proceedings.
- Posch, M., König, F., Branson, M., Brannath, W., Dunger-Baldauf, C., Bauer, P. (2005) Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine*, 24, 3697-3714.
- Schäfer, H., Müller, H. H. (2001). Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections.*Statistics in Medicine* 20: 3741–3751.
- Sellke, T., Siegmund, S. (1982). Sequential analysis of the proportional hazards model. *Biometrika* 70: 315–326.
- Shen, Y., Cai, J. (2003). Sample size re-estimation for clinical trials with censored survival data. *J. Amer. Stat. Ass.* 98: 418–426.
- Tsiatis, A. A. (1981). The asymptotic joint distribution of the efficient scores test for the proprtional hazards model calculated over time.
  Biometrika 68, 311–315.
- Tsiatis, A. A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. *J. Amer. Stat. Ass.* 77, 855–861.
- Wassmer, G. (2006). Planning and analyzing adaptive group sequential survival trials. *Biometrical J.* 48, 714-729.