CLINICAL INVESTIGATION

Adaptive designs for clinical trials with multiple endpoints

"Whenever several endpoints are included in the confirmatory analysis of a clinical trial … the calculation of an adequate sample size … is a particular challenge…"

Keywords: adaptive designs • clinical trials • multiple endpoints

Geraldine Rauch*,[1]
& Meinhard Kieser[1]
[1]Institute of Medical Biometry & Informatics, University of Heidelberg, Im Neuenheimer Feld 305, D-69120 Heidelberg, Germany
*Author for correspondence:
Tel.: +49 6221 561932
Fax: +49 6221 564195
rauch@imbi.uni-heidelberg.de

Rationale for adaptive designs with multiple endpoints

In clinical trials, the efficacy of a new treatment is usually assessed by a single primary endpoint [1,2]. Other important aspects concerning the new treatment, such as adverse events or quality of life, are taken into account by the definition of secondary endpoints. Usually, only the primary endpoint is analyzed confirmatorily with a statistical test whereas the secondary endpoints are evaluated with descriptive statistical methods. A descriptive analysis, however, does not result in a definite proof and therefore provides only limited additional information. The current research on multiple testing strategies offers many opportunities to incorporate secondary endpoints in the confirmatory analysis of a trial, for example, by a predefined hierarchical testing procedure or enhanced gatekeeping strategies [3–6]. While the nature of secondary endpoints is that they are only of subordinate clinical importance, there also exist many clinical trial applications where the efficacy of a new treatment cannot be adequately described by only one primary endpoint. The EMEA Guideline Points to Consider on Multiplicity Issues states that 'If, however, a single variable is not sufficient to capture the range of clinically relevant treatment benefits, the use of more than one primary variable may become necessary' [2]. In some situations, the efficacy claim is therefore based on the significance of several primary endpoints, also referred to as coprimary endpoints. However, even if several primary endpoints are considered, it is not always necessary to demonstrate a significant and relevant effect in all primary endpoints under investigation. Another option is to base the efficacy claim of the new intervention on the significance of at least one endpoint out of a predefined set of primary endpoints considered as clinically relevant.

Whenever several endpoints are included in the confirmatory analysis of a clinical trial, a multiple testing problem arises. In order to control the probability to falsely reject at least one of the null hypotheses under investigation at a predefined global significance level α, the local significance levels corresponding to the individual test hypotheses have to be adjusted by an adequate multiple testing procedure. There exist a variety of multiple testing procedures in the statistical literature for many different possible applications [3–8].

When several endpoints are assessed within a multiple test problem, the power of the trial depends on the expected effects of all endpoints and on the correlations between them. The calculation of an adequate sample size therefore is a particular challenge, as the number of required parameter assumptions to be estimated in the planning stage is much larger as compared with a single endpoint test problem. For this reason, adaptive study designs which allow reacting flexibly to these uncertainties in the planning stage play a major role for clinical trials with multiple endpoints.

Adaptive design strategies for multiple endpoints

A variety of adaptive trial designs have been proposed in the statistical literature addressing the different scenarios introduced above.

FUTURE SCIENCE part of fsg

When the aim is to incorporate secondary endpoints in the confirmatory analysis, a natural way would be to use the primary endpoint as a gatekeeper before testing the secondary endpoints. Such hierarchical testing strategies embedded in an adaptive design have been proposed by several authors [9–16]. Adaptive design strategies for multiple primary endpoints were addressed in [11,13,17–18].

The adaptive changes made during the interim analysis can be of various types such as sample size re-estimation [16], change of the testing hierarchy [12], change of the primary endpoint [17] or adjustment of stopping boundaries based on a recalculation of nuisance parameters, for example, the correlation between the test statistics [15].

Challenges of adaptive designs with multiple endpoints

Although most adaptive design strategies and corresponding stopping boundaries proposed for the classical case of one primary endpoint can be easily adapted to the setting of multiple endpoints [11,13], there also exist some particular challenges when several endpoints are considered.

First of all, an easy application of a standard adaptive design strategy to a multiple testing problem is only possible if all endpoints should be assessed at the same interim time points. In some clinical applications, however, it might seem more reasonable to evaluate some endpoints at an early stage and others later. Such approaches are also possible within an adaptive design, but require a careful and sound definition of the underlying test problem and the type I error probabilities to be controlled.

> "...adaptive study designs ... allow reacting flexibly to ... uncertainties in the planning stage..."

Second, the definition of optimal interim time points based on the assumed information fraction is not straightforward, especially if one or several time-to-event endpoints are considered. For continuous or binary endpoints, the information fraction corresponds to the fraction of patients recruited until the interim analysis with respect to the maximal number of patients to be recruited until the final analysis. For time-to-event data, however, the information fraction corresponds to the expected proportion of observed events at interim with respect to the total number of expected events at the final analysis [19]. Therefore, in case that at least one time-to-event endpoint is under investigation, the information fractions at a fixed interim time point usually deviate between the different endpoints. As a consequence, in the final analysis the data collected from the first and the second stage might not be weighted equally for both endpoints.

Finally, the test statistics corresponding to the different endpoints are often correlated. This correlation structure can be used to optimize the decision boundaries in order to increase the global power of the test procedure [10,15]. However, the underlying correlation matrix is hardly ever known in advance, so a recalculation of this nuisance parameter might be meaningful [15].

Open topics & further research

Due to the variety of possible clinical applications there also exists a broad range of possible extensions for adaptive designs with multiple endpoints.

An exemplary current topic is the combination of short-term endpoints and related long-term endpoints in seamless Phase II/II trials [20–22]. For this application, a high correlation between the surrogate or short-term endpoint and the Phase III endpoint is of major importance as a low correlation might question the adequacy of the surrogate.

Another important research field is the application of adaptive designs to several time-to-event endpoints [19]. Here, the timing of the interim analysis and the assumed information fractions at interim are of major importance, where the latter aspect has already been discussed above. With respect to the timing of the interim analysis, the information fraction at interim for both endpoints must be high enough to provide a least some evidence which supports a late interim analysis. On the contrary, an interim analysis is only meaningful during the regular recruitment Phase of the trial as a restart of patient recruitment after a recruitment stop plus a follow-up period is usually unrealistic in practice.

Of course, the different design strategies discussed in here and the variety of possible flexible changes during the interim analysis (estimation of nuisance parameters, sample size re-calculation, change of endpoints or change of the multiple testing strategy) might also be combined in various ways. Therefore, there remain a number of interesting topics for further research in this field.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1 ICH E9 Guideline. Statistical principles for clinical trials (1998). www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002928.pdf

2 CPMP Guideline. Points to consider on multiplicity issues in clinical trials (2002). www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf

3 Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeping multiple testing procedures. *J. Stat. Plan. Infer.* 99, 25–40 (2001).

4 Wiens BL. A fixed sequence Bonferroni procedure for testing multiple endpoints. *Pharm. Stat.* 2, 211–215 (2003).

5 Dmitrienko A, Offen WW, Westfall PH. Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Stat. Med.* 22, 2387–2400 (2003).

6 Chen X, Luo X, Caprizzi T. The application of enhanced gatekeeping strategies. *Stat. Med.* 24, 1385–1397 (2005).

7 Holm S. A simple sequentially directive multiple test procedure. *Scand. J. Stat.* 6, 65–70 (1979).

8 Schüler S, Mucha A, Doherty P, Kieser M, Rauch G. Easily applicable multiple testing procedures to improve the interpretation of clinical trials with composite endpoints. *Int. J. Cardiol.* 175, 126–132 (2014).

9 Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. *Stat. Med* 29, 219–228 (2010).

10 Hung HMJ, Wang SJ, O'Neill R. Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials *J. Biopharm. Stat.* 17, 6, 1201–1210 (2007).

11 Kieser M, Bauer P, Lehmacher W. Inference on multiple endpoints in clinical trials with adaptive interim analyses. *Biom. J.* 41, 261–277 (1999).

12 Kieser M. A note on adaptively changing the hierarchy of hypotheses in clinical trials with flexible design. *Drug Inf. J.* 39, 215–222 (2005).

13 Maurer W, Glimm E, Bretz F. Multiple and repeated testing of primary, coprimary, and secondary hypotheses. *Stat. Biopharm. Res.* 3, 336–352 (2011).

14 Tamhane AC, Mehta CR, Liu L. Testing a primary and a secondary endpoint in a group sequential design. *Biometrics* 66, 1174–1184 (2010).

15 Tamhane AC, Wu Y, Mehta CR. Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (I): unknown correlation between the endpoints. *Stat. Med.* 31, 2027–2040 (2012).

16 Tamhane AC, Wu Y, Mehta CR. Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (II): sample size re-estimation. *Stat. Med.* 31, 2041–2054 (2012).

17 Chen LM, Ibrahim JG, Chu H. Flexible stopping boundaries when changing primary endpoints after unblinded interim analyses. *J. Biopharm. Stat.* 24, 817–833 (2014).

18 Ye Y, Li A, Liu L, Yao B. A group sequential Holm procedure with multiple primary endpoints. *Stat. Med.* 32, 1112–1124 (2013).

19 Wassmer G. Planning and analyzing adaptive group-sequential survival trials. *Biom. J.* 48, 714–729 (2006).

20 Kunz, CU, Friede T, Parsons N, Todd S, Stallard N. Data-driven treatment selection for seamless Phase II/III trials incorporating early-outcome data. *Pharm. Stat.* 13, 238–246.

21 Stallard N. A confirmatory seamless Phase II/III clinical trial design incorporating short-term endpoint information. *Stat. Med.* 29, 959–971 (2010).

22 Todd S, Stallard N. A new clinical trial design combining Phases II and III: Sequential designs with treatment selection and a change of endpoint. *Drug Inf. J.* 39, 109–118 (2005).