Anja Sander
Dr. sc. hum.

Internal pilot study design in clinical trials with multiple binary endpoints

Fach: Medizinische Biometrie u. Informatik

Doktorvater: Prof. Dr. sc. hum. Meinhard Kieser

In clinical trials, it is generally recommended to define one primary endpoint which is then used to evaluate treatment effects and to proof efficacy. However, there are situations where it is difficult or not reasonable to select one single outcome as primary endpoint. Having defined multiple primary endpoints, this directly impacts the sample size calculation which is a key step in the planning phase and which requires the pre-specification of various parameters. Apart from the type I error rate, the target power, and the relevant treatment effect, nuisance parameters such as the variance in case of a continuous outcome and the overall event rate in the binary case must be specified. In the context of multiple endpoints, the correlations between the test statistics also determine the sample size necessary to reach a predefined power. Assumptions made in the planning stage are often related with uncertainty and therefore there is a high risk not to achieve the intended power. This uncertainty can be dealt with an internal pilot study design. In such a design, based on a first portion of patients nuisance parameters are estimated in a blinded way mid-course and the sample size is revised accordingly.

This work focused on the IPS design in the context of multiple primary binary endpoints. Existing methods for one binary endpoint were extended to three different testing situations.
First, clinical trials that aim at demonstrating a significant treatment effect for a composite endpoint or one pre-specified main component were addressed. Here, the Bonferroni correction was used to adjust for multiplicity with the aim to control the overall type I error rate. Since this correction method becomes very conservative with increasing correlation between the test statistics under the null hypothesis, the correlation was used to further increase the local alpha levels which leads to a better exhaustion of the global significance level.
Second, the general case of multiple primary binary endpoints was investigated comparing four different adjustment approaches: the Bonferroni correction, Holm's procedure, Hochberg's procedure, and the James test.
Third, the situation of binary co-primary endpoints with the aim to show an effect in all of them was evaluated. In this case, inflation of the overall type I error rate does not occur if all primary endpoints are tested at the global significance level.

The proposed IPS designs were described in detail, especially considering the mid-course estimation of correlations. Exact formulas on the actual type I error rate and power for the fixed sample size design and the IPS designs were provided for each of the considered situations for the case of two endpoints. For more than two endpoints, the notation becomes very complex but exact calculations are theoretically possible when this is of interest or required, for instance by regulatory agencies. Due to the high computational effort, the description of the performance characteristics of the IPS design were based on simulation studies by means of the actual type I error rate, power, and recalculated sample size within a broad range of settings.

The main results are as follows. The proposed IPS design turned out to be comparable with the fixed sample size design in view to the type I error rate or even better in terms of alpha exhaustion. Especially in case of mis-specification in nuisance parameters at the planning stage, the strength of the IPS design became obvious. At this, the estimation of the correlations is an improvement in the context of multiple endpoints. In general, as long as the IPS did not include already too many

patients, the intended power was achieved quite well. This holds true also in cases of relatively small sample sizes constituting the IPS. The recalculated sample size was in mean only a few patients different from that required in the fixed sample size design, but the smaller the IPS the higher the respective variability in the final sample size. However, depending on the size of the IPS and the sample size in general, bias in the event rates and correlations occurred due to the recalculation step but this is a general characteristic of such a design with sample size recalculation.

Clinical trial examples were used to illustrate the application of the different procedures. Moreover, R programs are provided which can be used to calculate correlation-adjusted local levels, to (re-)calculate sample sizes, and to calculate actual type I error rates and power, respectively.

This work provides evidence on the applicability of the internal pilot study design when multiple binary endpoints are considered. As the inclusion of an adequate number of patients is of main importance, implementation of the proposed designs in practice is desirable.