Discussion

Survival Endpoints and Adaptive Designs

Peter Bauer*

Core Unit for Medical Statistics and Informatics, Section of Medical Statistics,
Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria

Summary

This is a discussion of the paper 'Planning and analyzing adaptive group sequential survival trials' by Gernot Wassmer, appearing in this special issue on adaptive designs.

Key words: Adaptive design; Clinical trial; Group sequential test; Interim analysis; Log-rank test; Survival analysis.

Wassmer's paper (Wassmer, 2006) deals with adaptive design methodology. The author gives a framework for applying adaptive designs in studies with two treatment groups and a time till event endpoint. Using the asymptotic normal independent increment structure of stage-wise logrank statistics he can use the machinery of group sequential designs. Schoenfeld's estimate is used for the number of events required to achieve a certain conditional power, sample sizes are determined by assuming exponential survival and patient entering the trial at a constant rate during the accrual period. Also calendar times of the interim analyses (defined in fraction of the maximum information) are calculated. The inverse normal combination function is used for the adaptive test statistics where the weights remain fixed throughout the trial. (The option of the recursive combination test is not used but described at the end of Section 3.4). The paper is not the first one to demonstrate how to apply adaptive designs to time till event data. But it also proposes different estimates and confidence intervals for the hazard rate (constant by assumption): 1) The (conservative) repeated confidence interval for group sequential designs applied to the adaptive version of the test (RCI), which can be applied at any interim analysis. 2) The "monotone" confidence interval assumes an order in the sample space where earlier rejections are more extreme irrespective of the value of the test statistics. It can only be applied when a stopping criterion has been met or otherwise at the maximum sample size. For point estimation three options are discussed, a simple estimate using the exponential of the logrank statistics, the midpoint of the RCI or a median unbiased estimate defined as the limit of the one-sided monotone 50% confidence interval. Two overall p-values are proposed: One is defined as the lowest significance level so that the test (using the same family of rejection boundaries) would just reject at the given stage. The other can be calculated only after stopping and as usual denotes the probability of getting a more extreme outcome than observed (here based on the particular order in the sample space).

For sample size reassessment the number of events needed up to the next interim analysis in order to achieve a given conditional power for a rejection at the next stage is determined. The way to determine the number of events needed up to the end of the trial to achieve a certain overall conditional power is also described. Another option is to determine the number of events necessary to achieve a certain length of the RCI for the hazard ratio at the next interim analysis (leaving open conceptual questions on its usefulness). Simulations are performed to show the impact of constraint

* e-mail: peter.bauer@meduniwien.ac.at

sample size reassessment on power and average sample size, which show the known tendency of a rather large sample size close to the null hypothesis. The confidence level of the RCI is conservative, at the one hand by construction and at the other hand because of using approximations. These approximations are reasonable only if the hazard ratio is not too far from 1 and if the numbers of patients at risk in the treatment groups are roughly equal. They lead to an overestimation of the standard error for the hazard ratio, so that also the monotone confidence interval is conservative with a higher coverage probability than targeted. The author applied the methodology (which has been implemented in the software ADDPLAN) to compare the results with those from the literature. He also demonstrates how to proceed when an unscheduled interim analysis is performed before the first scheduled interim analysis, but the method does not seem to cover the option of calculating the conditional error at any unscheduled analysis time.

One of the main motivations behind the fine and hard piece of work Gernot Wassmer has done by implementing these tools in his software is that sample size reassessment "might help to rescue an underpowered trial". However, here I have some reservations about how sample size reassessment is quite frequently applied. One starts with a rather optimistic effect to get the project on track (for which the price might be that other more realistically designed projects will not be started). Resorting to the option of increasing the sample size during the trial may sometimes be used instead of a serious confrontation with the sample size issue in the planning phase.

In summary, Wassmer shows how flexibility can be applied to clinical trials with time till event endpoints. Even software to support the calculations is made available for this purpose.

References

Wassmer, G. (2006). Planning and analyzing adaptive group sequential survival trials. *Biometrical Journal* **48**, 714–729.

