Cornelia Ursula Kunz
Dr. sc. hum.

**Two-Stage Designs for Phase II Trials with One or Two Endpoints**

Promotionsfach: Medizinische Biometrie und Informatik
Doktorvater: Prof. Dr. sc. hum. M. Kieser

Cancer is still a leading cause of death worldwide and action has to be taken to prevent and to treat cancer. Within the development of new treatments, Phase II trials play a key role. The aim of these trials is to study the potential efficacy of the treatment under investigation which is assessed in terms of anti-cancer activity, for example, tumor shrinkage, with the two possible outcomes “yes” or “no”. For ethical reasons, an interim analysis is performed to allow an early termination of the trial if the treatment is inefficient. Although there is a current debate on whether those trials should be performed as randomized trials, the standard approach still consists of a single-arm trial.

Several approaches are available for the design of Phase II trials in oncology. The most commonly used design is probably Simon’s two-stage design. Various methodological extensions have been proposed to overcome some of the limitations of Simon’s design. One of these limitations is that Simon’s design is based on only one endpoint. In practice, it is often desirable to include further clinically relevant endpoints such as “survival after one year” to obtain more details about the mode of action of the new treatment. Although designs with two endpoints have been proposed in the literature, there is still a need for new methods. The existing designs are restricted to rather specific situations, such as specific assumptions about the structure of the endpoints considered, as for example, the outcomes of one endpoint are a subset of the outcomes of the other endpoint. Furthermore, while methods exist to calculate the sample size for some of those trials, little research has been done regarding the analysis of the obtained data. Until now, the usual analysis of the data is done using methods derived for one-stage designs. Those methods lead to biased results since they do not take into account the sequential design of the study.

Within this thesis, methods are derived to overcome some of the limitations of the existing approaches. The methods include analytical solutions for sample size calculations and the construction of a closed test procedure allowing inference to be made not only about the global null hypothesis but also about the single hypotheses for the two considered endpoints. In addition, point estimators, confidence sets, and confidence intervals are derived that are appropriate to analyze the obtained data from two-stage designs. Furthermore, to allow for early stopping whenever it becomes evident that the trial will not conclude with a significant result, curtailment rules are implemented in the designs and the effect of those rules is studied. Examples are used to illustrate the application of the methods.

For Simon’s design, the effect of implementing curtailment rules have been studied within this thesis. Analytical solutions regarding type I error rate, power, expected sample size and probability of early termination could be derived for the implementation of non-stochastic curtailment. The effect of stochastic curtailment has been investigated using simulation studies. While non-stochastic curtailed designs always have the same type I error rate and the same power as uncurtailed designs, the savings in total sample size can be large (up to around 40%). Stochastic curtailment leads to even higher savings in sample size but those designs also have a smaller power compared to the uncurtailed designs. Overall, the effect of implementing curtailment rules is highly dependent on the specific characteristics of the design.

Special focus was put on the design with two endpoints where the outcomes of one endpoint are a subset of the outcomes of the other endpoint. The analytical solution to calculate the critical values and sample sizes for this design is given. Results are tabulated for a wide range of parameter values. A comparison with Simon's design showed that using the design with two endpoints can lead to savings in total and expected sample size. Furthermore, nearly all tabulated examples also provide a closed test procedure. In addition to the analytical solution to determine critical values and sample sizes, the maximum likelihood estimators for the true response rates were derived. Although the solutions are rather intuitive, these estimators are biased and the true response rates are always underestimated. A bias of up to 4% was found for the investigated examples. Furthermore, bias-reduced estimators were also investigated. Although these estimators are less biased, the response rates can now be over- and underestimated. Therefore, an unbiased estimator is derived to obtain correct estimates of the true response rate. The disadvantage of the unbiased estimator is that the calculation involved is rather complex. Since it is common practice not only to report p-values and point estimators, confidence sets for the global null hypothesis as well as confidence intervals for the single hypotheses are given. The interpretation of the resulting confidence sets and confidence intervals are consistent with the hypothesis test. Finally, curtailment rules were implemented in this design. As for Simon's design, the effect of curtailment is highly dependent on the characteristics of the uncurtailed designs. However, savings in expected sample size can be achieved, in some cases with no or only small loss in power. For the other designs with two endpoints, analytical solutions to calculate the critical values and sample sizes are given. Solutions are tabulated for wide ranges of parameter values. For the investigated examples, the total and expected sample sizes were always smaller than those for Simon's design. In the extreme case, the total sample size for the design with two endpoints was 62 less than the total sample size for Simon's design. However, many of those designs do not provide a closed test procedure to test the single endpoints. And even if they do, it does not necessarily mean that the power to detect an effect for the single endpoint is as high as the power for the global null hypothesis.

Some calculations involved in the methods considered within this thesis are computationally intensive. Therefore, ways to improve the runtime of the algorithms are developed and the implementation of the algorithms is described in detail. All programs are provided.

Overall, designs with two endpoints provide a useful alternative to Simon's design. Based on the time course and the importance of the endpoints, an appropriate design with two endpoints should be chosen and the characteristics of the design should be compared to those of Simon's design. In addition, the effect of implementing curtailment should be investigated and, if feasible, a curtailed design should be used. For data analysis, appropriate methods taking into account the two stages of the design, should be used.