CLINICAL INVESTIGATION

Adaptive designs for Phase II oncology trials

"In case that the interim results or external information suggest a change of the initially specified sample size for stage two, such an adaption can be performed while still assuring control of the type I error rate..."

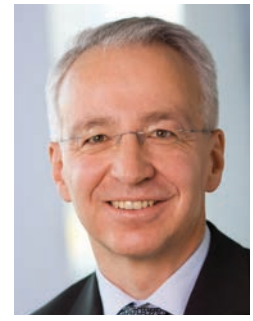
Keywords: adaptive designs • clinical trials • oncology drug development • Phase II

The role of Phase II trials in oncology drug development

Traditionally, Phase II in oncology drug development acts as a screening tool by identifying treatments with sufficient activity that warrant further investigation in large and costly Phase III studies. However, Phase II has also another more prominent role in cancer research as there is a significant number of drugs that were approved based on data of Phase II trials only. For example, according to the registration information for US FDA oncology drug approvals for solid tumours, for 17% of approvals between 1998 and 2008 Phase II data were the only basis [1]. Such a situation arises, for example, in case of an overwhelmingly positive result in Phase II, where a proceeding Phase III study might be impossible to perform due to ethical reasons. In the era of targeted therapies, this aspect may become of even growing importance. There is a continuing and lively debate about the merits of single-armed versus randomized Phase II trials in oncology [2–4]. It is acknowledged that 'efficient drug development will require the appropriate use of both single-arm and randomized Phase II trials' [2]. Even more, in view of the high activity demonstrated for some targeted drugs as compared with available conventional therapies, use of single-arm trials for drugs with exceptional early activity was recently promoted from regulatory side [5]. This viewpoint is supported by the finding that approvals based on nonrandomized trials with definite end points show a reassuring record of long-term safety and efficacy [6]. However, it is of utmost importance that efficient and valid study designs are applied to successfully and adequately address these aims of Phase II cancer trials.

'Classical' single-arm two-stage designs

In single-arm Phase II oncology studies, a binary end point is usually applied as primary outcome. In the past, this end point was usually a short-term or intermediate outcome related to tumor shrinkage. However, as early tumor shrinkage is not always related to extended survival – which is the ultimate goal of any cancer therapy – and as efficacious novel molecular targeted drugs may not show an early tumor shrinkage, alternative end points such as progression-free survival at a defined time point after start of treatment are increasingly applied. For ethical reasons and to speed up the drug development program, these trials are generally performed with an interim analysis to allow for early stopping for futility or efficacy. The null hypothesis of insufficient efficacy is tested by defining an uninteresting event rate that has to be outperformed by the new drug to demonstrate sufficient activity. Furthermore, specification of an anticipated event rate is required which expresses the expected efficacy being high enough to define a clinically relevant improvement. For these rates, the sample sizes of the two stages as well as the related decision rules for interim and final analyses can be determined such that the constraints with respect to type I error rate (i.e., probability of erroneously concluding that an ineffective agent is effective) and power (i.e., probability of demonstrating efficacy



Meinhard Kieser
Author for correspondence:
Institute of Medical Biometry & Informatics, University of Heidelberg, Im Neuenheimer Feld 305, D-69120 Heidelberg, Germany
Tel.: +49 6221 564140
Fax: +49 6221 564195
meinhard.kieser@imbi.uni-heidelberg.de

Stefan Englert
Institute of Medical Biometry & Informatics, University of Heidelberg, Im Neuenheimer Feld 305, D-69120 Heidelberg, Germany

Geraldine Rauch
Institute of Medical Biometry & Informatics, University of Heidelberg, Im Neuenheimer Feld 305, D-69120 Heidelberg, Germany

FUTURE SCIENCE part of fsg

for an effective agent) are fulfilled. Simon's design [7] is the most frequently used single-arm Phase II design in oncology drug development [8]. This design minimizes the maximum (minimax design) or expected sample size under the null hypothesis (optimal design) among all designs with the same significance level or power, and it includes the possibility to stop for futility after the first stage. Various extensions of Simon's two-stage design exist, for example minimization of sample size with respect to alternative criteria [9–11], designs with more than one interim analysis [12] or with the option of stopping both for futility and efficacy [13–15]. It is a common feature of all these designs that the sample size for the study stages as well as the rules for early stopping and rejection of the null hypothesis have to be defined in the protocol and have to be adhered to strictly when conducting and analyzing the trial. Otherwise, control of the significance level is no longer assured [16], which is essential, particularly when a Phase II study is the basis of approval. However, especially in early phases of drug development there is usually a considerable amount of uncertainty when planning a trial. For example, if the interim results strongly suggest that the activity of the drug under investigation is higher than assumed in the planning stage, the initial sample size considerations may become questionable, and without an adjustment of the sample size the study may be overpowered. As a further example, it is usually difficult to stop patient recruitment exactly when the number of patients to be included in the interim or final analysis is achieved. As a consequence, over- or under-running may occur which, however, cannot be adequately handled in the framework of 'classical' two-stage designs. Therefore, there is the need for study designs that allow flexible changes of design characteristics midcourse and that assure at the same time control of the type I error rate.

Adaptive single-arm two-stage designs

A decade ago, designs were proposed where the final sample size of the second study stage is determined based on the results observed in the interim analysis [17,18]. Thus, these approaches do not exclusively rely on the original planning assumptions when fixing the trial design but also take into account information accrued during the course of trial. However, it is a disadvantage of these methods that the rules for sample size adaption have to be prespecified in the trial protocol. By this, these designs are not flexible enough to enable an adequate handling of unforeseeable occurrences. There are a multitude of aspects that may contribute to the decision of modifying a design, not only the interim results of the current trial but also external information on the same or alternative drugs. Such a more far-reaching flexibility can be achieved by applying the so-called discrete conditional error function methodology [19,20]. Roughly speaking, the conditional error function defines the significance level to be used for the second study stage depending on the results of the interim analysis. If this function is defined appropriately and the level for stage two is chosen accordingly, the overall significance level is controlled even if a (data-dependent) modification of the sample size was performed after the first stage [21,22]. This concept was originally proposed for controlled studies with continuous end points but could be adapted to single-arm two-stage designs taking into account the discrete nature of the outcome variable [19]. It was shown that any 'classical' two-stage design has a unique equivalent discrete conditional error function representation. As a consequence, using this approach enables to start a Phase II trial with a 'classical' design, for example with Simon's optimal design. In case that the interim results or external information suggest a change of the initially specified sample size for stage two, such an adaption can be performed while still assuring control of the type I error rate; otherwise the initially specified design is maintained. Note that due to this flexibility with respect to the second stage sample size, the method includes as a by-product also control of the type I error rate in case of over- or under-running the sample size envisaged for the final analysis. Furthermore, it turns out that the 'classical' two-stage designs are a proper subset of those designs that can be characterized by a discrete conditional error function. Exploiting the more general discrete conditional error function approach, designs could be identified that are more efficient in terms of expected or maximum sample size, respectively, as compared with their 'classical' counterparts while being at the same time more flexible [19].

Illustrative example

To illustrate the opportunities of adaptive designs in Phase II oncology trials, we consider a situation similar to the one described in [23] where a new treatment option for patient with pancreatic neuroendocrine tumors was investigated. For the end point 6-month progression-free survival, an uninteresting rate of 60% obtained from historical controls and an anticipated rate of 80% were assumed. Simon's optimal design for a one-sided significance level of 0.05 and a power of 0.90 includes 19 patients in the first stage. If more than 12 of these patients are progression-free after 6 months, the trial continues with further 34 patients. In the final analysis, the null hypothesis is rejected if more than 37 of the total number of 53 patients are progression-free after 6 months. Let us now assume that 17 (89%) of the 19 patients analyzed in the interim analysis were

progression-free. In the classical approach, further 34 patients have to be recruited for the second stage although only 21 (62%) further patients without progression after 6 months are required to demonstrate efficacy. Expressed equivalently in terms of the related discrete conditional error function, the null hypothesis can be rejected if the p-value calculated from the second stage data falls below 0.4908. For the original design, conditional on the interim result the statistical power for a true rate of 80% amounts to 99.6% and even amounts to 99.9% if the true rate of progression-free survival equals to the rate observed in the interim analysis. Thus, the interim results seem to indicate that a reasonable power may be realized with a much lower sample size than initially planned. In an adaptive design, the sample size of the second stage can be recalculated. Technically, the calculation is performed as for a single-stage binomial test where the significance level is set equal to 0.4908 as defined by the discrete conditional error function. Assuming a true rate of 80% (as originally planned) or 89% (as observed in the interim analysis), additional 9 or 4 patients, respectively, are sufficient to achieve 90% power. The adaptive design allows switching to any of these sample sizes while still controlling the type I error rate.

Further results & future research topics

The increase in flexibility of study designs is both a blessing and a curse. On the one hand, it enables to incorporate knowledge gained after the planning stage when choosing the definite design, thus allowing to learn from accruing information. On the other hand, when the decision to modify an initially specified design is data-driven, there is the risk to be fooled by the data as they are subject to chance due to their inherent variability. Therefore, the question arises which design modifications are efficient in which situations. For the topic of sample size recalculation, a comprehensive investigation was performed that investigated the performance of various recalculation strategies and whose results can be used to identify the most appropriate approach for specific trial situation at hand [24]. It was already mentioned that the conditional error function approach enables to deal with over- or under-running the sample size targeted for the final analysis in a simple and elegant way: as the decision rule for rejecting or accepting the null hypothesis is defined in terms of the p-value calculated for the second stage data and as this quantity solely depends on the prespecified conditional error function and the interim results, using this approach guarantees that the significance level is controlled for arbitrary sample size of the second stage. As a further step forward, a general methodology could be developed that also allows handling both unintentional and intentional over- or under-running in the first stage while strictly controlling the type I error rate [25].

Major progress has been made in the past few years to develop adaptive design methodology for Phase II oncology trials, but there are still a number of open issues that are currently under research. For example, 'classical' two-stage designs have been proposed recently for single-armed cancer trials where the end point is time to the occurrence of a defined event, for example disease progression or death [26]. Flexible counterparts to these designs would be highly attractive. Finally, there are clinical trial situations, where two primary end points are considered [27]. For example, the study mentioned above [23] included both objective tumor response and 6-month progression-free survival as primary outcome variables. 'Classical' optimal designs have been derived [28] for these applications, but a related adaptive framework has still to be developed.

Financial & competing interests disclosure

The authors have no other 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 apart from those disclosed.

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

References

1 US Food and Drug Administration: Drugs@FDA. www.accessdata.fda.gov

2 Gan HK, Grothey A, Pond GR, Moore MJ, Siu LL, Sargent D. Randomized Phase II trials: inevitable or inadvisable? *J. Clin. Oncol.* 28(15), 2641–2647 (2010)

3 Stewart DJ. Randomized Phase II trials: misleading and unreliable. *J. Clin. Oncol.* 28(31), e649–e650; author reply e651–e653 (2010).

4 Pond GR, Abbasi S. Quantitative evaluation of single-arm versus randomized Phase II cancer clinical trials. *Clin. Trials* 8(3), 260–269 (2011).

5 US FDA. The next phase in oncology: FDA's Pazdur has new vision for drug development. *The Pink Sheet*, 11th November (2013).

6 Tsimberidou AM, Braiteh F, Stewart DJ, Kurzrock R. Ultimate fate of oncology drugs approved by the us food and drug administration without a randomized trial. *J. Clin. Oncol.* 27(36), 6243–6250 (2009).

7 Simon R. Optimal two-stage designs for Phase II clinical trials. *Control. Clin. Trials* 10, 1–10 (1989).

8 Thezenas S, Duffour J, Culine S, Kramar A. Five-year change in statistical designs of Phase II trials published in leading cancer journals. *Eur. J. Cancer* 40(8), 1244–1249 (2004).

9 Jung SH, Lee T, Kim K, George SL. Admissible two-stage designs for Phase II cancer clinical trials. *Stat. Med.* 23(4), 561–569 (2004).

10 Shuster J. Optimal two-stage designs for single arm Phase II cancer trials. *J. Biopharm. Stat.* 12(1), 39–51 (2002).

11 Mander AP, Wason JMS, Sweeting MJ, Thompson SG. Admissible two-stage designs for Phase II cancer clinical trials that incorporate the expected sample size under the alternative hypothesis. *Pharm. Stat.* 11(2), 91–96 (2012).

12 Chen TT. Optimal three-stage designs for Phase II cancer trials. *Stat. Med.* 16, 2701–2711 (1997).

13 Fleming TR. One-sample multiple testing procedure for Phase II clinical trials. *Biometrics* 38, 143–151 (1982).

14 Chang MN, Therneau TM, Wieand HS, Cha SS. Designs for group sequential Phase II clinical trials. *Biometrics* 43(4), 865–874 (1987).

15 Mander AP, Thompson SG. Two-stage designs optimal under the alternative hypothesis for Phase II cancer clinical trials. *Contemp. Clin. Trials* 31(6), 572–578 (2010).

16 Englert S, Kieser M. Adaptive designs for single-arm Phase II trials in oncology. *Pharm. Stat.* 11(3), 241–249 (2012).

17 Lin Y, Shih WJ. Adaptive two-stage designs for single-arm Phase IIa cancer clinical trials. *Biometrics* 60, 482–490 (2004).

18 Banerjee A, Tsiatis AA. Adaptive two-stage designs in Phase II clinical trials. *Stat. Med.* 25(19), 3382–3395 (2006).

19 Englert S, Kieser M. Improving the flexibility and efficiency of Phase II designs for oncology trials. *Biometrics* 68(3), 886–892 (2012).

20 Jin H, Wei Z. A new adaptive design based on Simon's two-stage optimal design for Phase II clinical trials. *Contemp. Clin. Trials* 33(6), 1255–1260 (2012).

21 Proschan M, Hunsberger S. Designed extension of studies based on conditional power. *Biometrics* 51, 1315–1324 (1995).

22 Müller H, Schäfer H. Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. *Biometrics* 57, 886–891 (2001).

23 Hobday TJ, Qin R, Reidy-Lagunes D *et al*. Multicenter Phase II trial of temsirolimus and bevacizumab in pancreatic neuroendocrine tumors. *J. Clin. Oncol.* pii: JCO.2014.56.2082 (2014) (Epub ahead of print).

24 Kieser M, Englert S. Performance of adaptive designs for single-armed Phase II oncology trials. *J. Biopharm. Stat.* (2014) (Epub ahead of print).

25 Englert S, Kieser M. Methods for proper handling of over- and underrunning in Phase II designs for oncology trials. Technical Report. *Stat. Med.* (2015) (In Press).

26 Kwak M, Jung SH. Phase II clinical trials with time-to-event endpoints: optimal two-stage designs with one-sample log-rank test. *Stat. Med.* 33(12), 2004–2016 (2014).

27 Lin X, Allred R, Andrews G. A two-stage Phase II trial design utilizing both primary and secondary endpoints. *Pharm. Stat.* 7(2), 88–92 (2008).

28 Kunz CU, Kieser M. Optimal two-stage designs for single-arm Phase II oncology trials with two binary endpoints. *Methods Inf. Med.* 50(4), 372–377 (2011).