The adaptive design: state of the art and paths for future research

Raphaël Porcher

Department of Biostatistics and Medical Informatics, Saint-Louis Hospital,
Paris Diderot University, Paris, France

Corresponding Author:
Raphaël Porcher, Département de Biostatistique et Informatique Médicale,
Hôpital Saint-Louis, 1 avenue Claude Vellefaux, 75475 Paris cedex 10, France
E-mail: raphael.porcher@univ-paris-diderot.fr

Summary
The term adaptive design refers to situations in which accumulated data are used to modify some aspects of an ongoing clinical trial. Methods for adaptive designs aim to preserve the validity and the integrity of the trial despite the design modifications. Following the publication of FDA reports stressing the need to develop new tools to enhance the process of drug development, much interest and debate on adaptive trials has appeared in the literature.
In this paper, we focus on the principles and methodology of flexible trials, i.e. group sequential trials in which design modifications are allowed at interim analysis.

KEY WORDS: *adaptive designs, group-sequential trials, flexible designs, mid-course modifications, sample size reassessment.*

Introduction

Adaptive designs have received much attention in the recent literature, both from a methodological and a practical point of view [e.g. the white paper from the PhRMA working group (1), special issues of the *Journal of Biopharmaceutical Statistics* in 2005 and of the *Biometrical Journal* in 2006, a *JAMA* editorial in 2006 (2), ...]. While the methodological literature on adaptive designs was already well developed in statistical journals after the publication of seminal papers in the '90s (3-7), more recently, much debate and discussion has followed the publication of FDA reports urging progress in developing and using new tools to improve the efficacy of drug development (8-11). Interestingly, these discussions have involved statisticians and trialists from the pharmaceutical industry, the academic world, and regulatory agencies.
As opposed to traditional fixed design trials, adaptive trials are clinical trials where accumulated data are used to modify some aspects of the ongoing trial, while still preserving its validity and integrity. This definition encompasses a broad range of possibilities. For instance, group sequential designs may, in a way, be considered adaptive designs, as stopping rules allow early termination of the trial on the basis of the evidence accumulated in the course of it. Similarly, trials where the treatment allocation of new patients depends on the outcomes of previous patients, as in the case of the play-the-winner adaptive allocation scheme, are also adaptive trials (12). Adaptive trials may also comprise Bayesian single-arm or multi-arm designs (13). One important point in order to understand adaptive designs is that adaptation is a design feature aiming at enhancing the trial, and not simply ad hoc changes used as a way of rescuing a poorly planned trial (1).
Since the field covered by adaptive designs is quite wide, this paper will focus more specifically on so-called *flexible* designs, i.e. group sequential trials in which design modifications are allowed at interim analysis, using data from inside and outside the trial (14-16). The paper is organised as follows. The second section describes the principles of flexible designs and some types of adaptation already used in the literature, and presents seamless phase II/III trials within the framework of flexible trials. The third section gives an outline of statistical methods used to

control the type I error rate of a flexible trial, and briefly underlines current methodological issues. The last section contains discussion on flexible trials and future research.

Principles of flexible designs

Flexible designs are mainly an extension of group sequential designs, where the sequel of the trial may be re-designed after a planned or unplanned interim analysis. One principal justification for flexible designs is their ability to cope with some of the drawbacks of group sequential trials, which rely on the specification a priori of key parameters, such as the expected treatment effect one wishes to detect, the variance of the main endpoint, the treatment arms to be compared, and the statistical methods used to carry out these comparisons. The predetermination of all these characteristics may not always be an easy task, and misspecification of them can occur in inefficient trials. In contrast, a flexible design may make it possible to reassess the sample size of the remainder of the trial on the basis of the data of the trial itself, blinded or unblinded, or of information external to the trial. One may also choose a test statistic more appropriate to the trial data or modify a multiple testing procedure. In any case, the overall type I error rate of the trial has to be preserved. Control of the type I error is possible through adherence to an invariance principle. Consider that, without any design modification, the final test decision is based upon a combination of test statistics obtained at different stages of the trial, controlling for the type I error rate α. Any design modification which preserves the distributional properties of these stage-wise test statistics preserves α in the whole (14, 17). It thus becomes possible to replace the remainder of a trial with a design which preserves the initial type I error, conditional on what has been observed up to the intermediate analysis.

Flexible two-stage design

For the sake of simplicity, the statistical principle of flexible designs is presented for a two-stage design, though it can easily be extended to more than two stages. The sequential nature of flexible designs leads to the need to integrate information from the different stages into a single stage-wise test statistic. Several methods have been presented, which are easily explained in terms of combination tests (18, 19). Let us assume the trial is designed to test a null hypothesis H_{01}. The scheme of the two-stage design is presented Figure 1. To test the null hypothesis, a first stage is planned as in a conventional group sequential design. Type I (α) and II (β) error rates are fixed, and a test statistic Z is defined appropriately – we will here use the one-sided test p-value. Stopping limits for stage 1, namely α_0 ($> \alpha$) and α_1 ($< \alpha$), are determined, and a sample size n_1 is computed [see e.g. (20)]. At this stage, the way to combine the

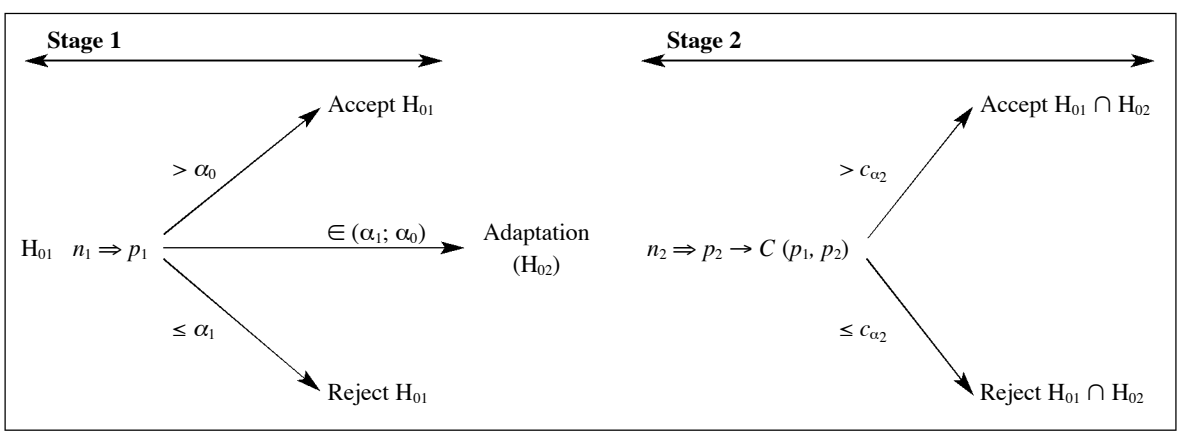

Figure 1. Schematic representation of a two-stage flexible design. The first stage allows testing of the null hypothesis H_{01}; at the second stage, a design adaptation is performed, possibly leading to selecting a second null hypothesis H_{02}. If no adaptation of the null hypothesis is decided, then H_{02} is simply H_{01}. Stage-wise p-values are combined as $C(p_1, p_2)$, and the final decision concerns $H_{01} \cap H_{02}$.

stage-wise test statistics has to be laid down. This combination rule has to be fixed before the interim analysis is carried out, but the stopping limits may be adapted at the time of the interim analysis (20, 21).

Once the n_1 subjects have been accrued, the test is performed, and a *p*-value p_1 is obtained. According to the value of p_1, the trial may stop according to the predefined stopping rules. If $p_1 \leq \alpha_1$, the trial stops with rejection of the null hypothesis (stopping for efficacy), and if $p_1 \geq \alpha_0$, the trial stops without rejection of H_{01}, which is also referred to as stopping for futility. If no early stopping rule is met, the trial proceeds to a second stage. Contrary to classic group sequential trials, it is, however, possible to use all available information to redesign the second stage (design adaptation). Design modifications can be based on data collected in the trial so far, but also on information external to the trial, and do not need to be laid down in advance (15), although it has been recommended to prespecify the scope of possible adaptations in order to facilitate interpretation and acceptance of the trial results (1). One common adaptation is to reassess the sample size n_2 to be recruited in stage 2.

The second stage of the trial is then conducted up to the accrual of n_2 patients, resulting in p_2, independent from p_1 under the global null hypothesis. Both *p*-values are combined according to the pre-specified rule $C(p_1,p_2)$, leading to rejection of the intersection null hypothesis $H_{01} \cap H_{02}$, or not, according to whether or not $C(p_1,p_2) \leq c_{\alpha_2}$. The critical value c_{α_2} depends on the combination function, and on the early rejection limits α_0 and α_1. It is chosen to control the overall type I error rate α. Note that special cases are obtained setting $\alpha_0 = 1$ to avoid early stopping for futility, or $\alpha_1 = 0$ to avoid early rejection.

At interim analysis, one may also set two decision limits for the second stage, as for the first stage, and let the trial proceed to a third stage if $C(p_1,p_2)$ falls between the boundaries, and so on ... (19).

Adaptations

In the literature, most design adaptations considered so far have dealt with sample size reassessment (4, 5, 7, 20). First, sample size reassessment based on blinded or unblinded nuisance parameter estimates (e.g. variance) has been widely discussed, and is no longer a matter of debate, except with regard to unblinding (22-24). These design modifications principally aim to preserve the power of the trial in the event of misspecification of these nuisance parameters at the time of trial planning.

More debated is the use of the observed treatment effect in the sample size reassessment (25). The conditional power, i.e. the power conditional to the observed outcome at the interim analysis, seems a reasonable choice to recompute the sample size for the second stage. Actually, the overall power averages out over all possible outcomes at the interim analysis. Once this analysis is performed, there is no advantage in taking an expectation over what is known not to have occurred. Actually, conditional power was regarded as a tool for trial monitoring long before the spread of the literature on adaptive designs (26). The main question is under which alternative the conditional should be computed. Some authors have used the estimated treatment effect at the interim analysis, also called the predictive power (20, 27). If using this observed treatment effect provides a good control of power it may also lead to dramatically increased sample sizes (21, 25, 28). The experimenter must bear in mind that the observed treatment effect is a random variable, and careless plug-in of apparent large or small effects may have drastic consequences in terms of sample size. One alternative would be not to decrease the pre-planned sample size, but rather to insert an additional interim analysis (see below), and to set an upper limit to the sample size increase.

To gain advantage from interim analyses, some authors have also proposed a general framework to incorporate adaptive data-driven changes in sample size, number and timing of interim analyses, as well as alpha-spending function (16, 29). This framework allows design modifications at any time, and not only at pre-specified times. One may add an interim analysis if unblinded data indicate a high chance of obtaining a decision, or drop the next interim analysis if they do not. The methods are, however, rather complex and beyond the scope of this article.

Other adaptations considered in the literature concerned redefining a dose range (30), dropping or adding treatment arms (31, 32), or selecting a more appropriate test statistic, for example in cases of non-proportional hazards for survival data or to adjust a

test on a variable that was not part of the original protocol (33, 34). Another proposed adaptation is to select, add or even change endpoints (35, 36). While it may be conceivable to discard one of the endpoints to be combined in a composite score on the basis of poor reproducibility, for example, or to modify the hierarchy of a multiple testing procedure, it is difficult to advise changing an endpoint. Modifying the main endpoint of a trial will almost inevitably cast reasonable doubt on the trial conclusions, and obviously indicate poor trial planning.

This list of adaptations is not exhaustive, and any relevant design modification may be considered. The critical point is to preserve the trial's integrity and credibility. In particular, one should be careful when using parameters estimated at an interim analysis, as the main risk is that of modifying a good protocol on the basis of results observed on a possibly small number of patients. Also, as a combination test allows conclusion of the intersection null hypothesis, one has to verify that such a hypothesis makes sense before adapting it. Nevertheless, protocol amendments are common practice, including dropping treatment arms that are considered ineffective after an interim analysis or on the basis of external evidence, such as other clinical trials. In this respect, flexible designs at least provide a framework for designing adaptations with control of the overall type I error rate.

Seamless phase II/III trials

Besides design modifications such as sample size reassessment, seamless phase II/III clinical trials are among the most promising implementations of flexible designs (1).

Traditionally, drug development consists of first conducting a phase II trial, where the experimental treatment is evaluated in terms of short-term safety and efficacy. For this purpose, several doses are often compared to a control. When the phase II trial is completed, its results are used to take a decision on whether the drug development may be continued, and which doses, targeted treatment effect, and so on, should be used for a subsequent phase III trial. The final analysis of efficacy relies on a long-term clinical endpoint measured in the phase III trial only, ignoring the information from previous phase trials.

Seamless phase II/III designs combine both a phase II and a phase III into a single uninterrupted two-stage study (37, 38). Figure 2 gives one possible schematic representation of such a design. The first stage corresponds to a phase II trial, where an early endpoint is used for decision-making concerning the second stage (adaptation). For instance, one may use the results of the interim analysis to decide whether it is worth conducting a second stage, to select the doses or a targeted subpopulation, for example, the treatment effect, and the sample size for the second stage. After design adaptation, the second stage corresponds to a phase III trial on the selected doses, with a terminal endpoint. However, patients accrued during the first-stage are still followed up, allowing the evaluation of the terminal endpoint on the first-stage sample. Some authors have even proposed planning the second stage itself as a flexible trial, with, for example, sample size reassessment on the basis of the results of long-term follow up of the phase II patients (37). The final analysis thus includes patients from both stages, and is performed in order to control the overall type I error rate at a prespecified level regardless of the adaptations performed at the interim analysis, as in other flexible designs described in previous sections. The use of well-conducted seamless phase II/III designs offers several advantages. First, it allows reduction of the time of drug development, by shortening the interval between the separate phases. As the study is initially planned as a whole, the time for obtaining approval from institutional review boards may be shortened, although adaptations also need approval. Moreover, there is no interruption of patient enrolment. Second, it increases the value of the information contributed by the phase II study, it needs fewer patients, and it cuts costs by integrating the evidence from both stages. Last, long-term safety data may be available earlier, resulting from additional follow up of phase II patients.

Such combined phase II/III trials are confirmatory trials and are a sensible way to meet the expectations of pharmaceutical companies, health authorities, physicians, but above all patients, with regard to the possibility of making effective treatments available as early as possible. But planning and implementing

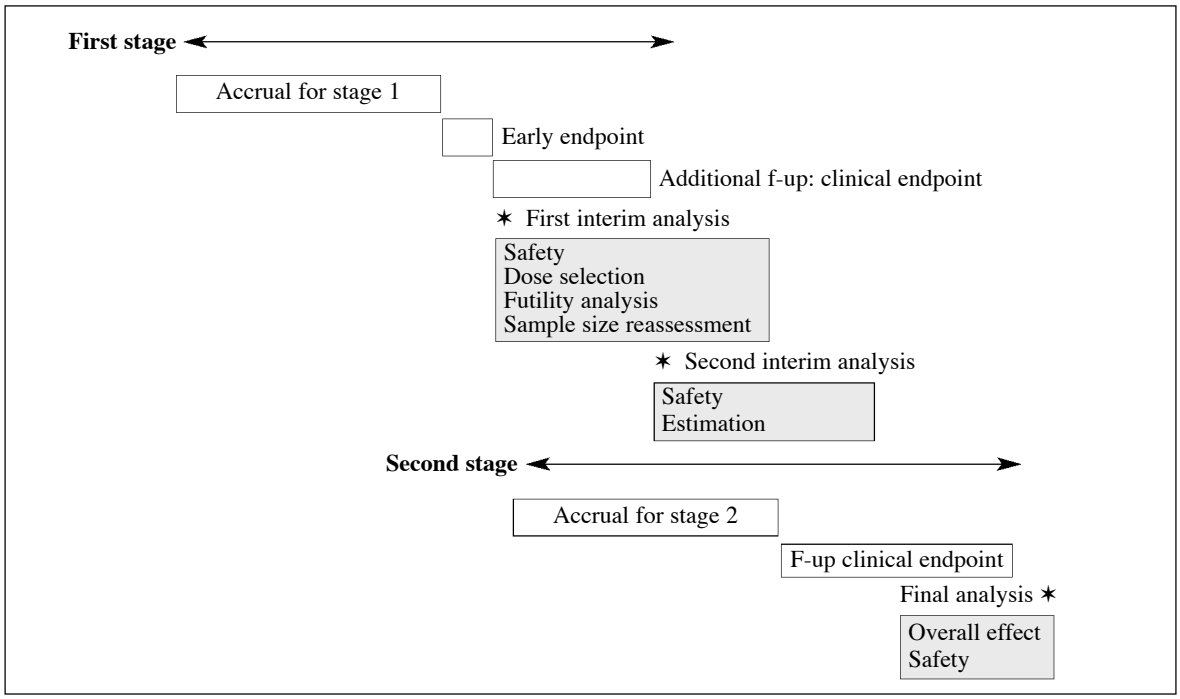

Figure 2. Schematic representation of a phase II/III seamless trial. The first stage corresponds to phase II, with additional follow up also used at the second stage, which corresponds to phase III. Stars (✶) denote times of analysis and grey hatched boxes the analyses that may be performed.

seamless phase II/III trials is also a challenge for statisticians and clinical teams. While the theory allows control of the overall type I error for a wide range of design adaptations, it remains important to assess the impact of these adaptations on the treatment effect estimates and on the operational characteristics of the trial (39).

Statistical methods and properties

In this section, we briefly give some insight into the statistical methods used to implement flexible designs, and discuss some issues concerning estimation following a flexible trial.

Statistical approaches for flexible designs

When introducing the principles of flexible designs in section 2, we relied on the notion of combination tests. Early works on flexible designs used combination of *p*-values (4), while others have proposed combination of normal deviates rather than *p*-values (7). Even though these practical methods are different, both are combination tests. Another method rests on the conditional error function (5), which, however, is also equivalent to a combination test (40). Of note, all these test statistics are different from the usual cumulative sequential test statistics used in group sequential trials.

Combining p-values

Recursive combination tests have been quite extensively studied by Brannath et al. (19). To combine *p*-values, the combination function $C(p_1,p_2)$ has to be increasing in both arguments, strictly increasing in at least one, and left-continuous in p_2 for $p_1 \in (\alpha_1;\alpha_0)$ and $p_2 \in [0; 1]$. Moreover, p_1 and p_2 have to be *p*-clud, i.e. $\Pr_{H_0}(p_1 \leq \alpha) \leq \alpha$ and $\Pr_{H_0}(p_2 \leq \alpha | p_1) \leq \alpha$, $\forall \alpha \in [0, 1]$. In practice, the *p*-clud condition may not be easy to determine, but independent and uniformly distributed p_1 and p_2 is a sufficient condition to obtain it.

Several combination functions may be considered. In their initial report, Bauer and Köhne used Fisher's product criterion $C(p_1,p_2) = p_1 \times p_2$, and this choice has been one of the most frequent in the subsequent literature (4). Recently, a sum of *p*-values instead of

a product has been considered (41). For Fisher's combination test, the critical value for the combination test is easily obtained using

$$p_1 p_2 \leq c_{\alpha_2} = \exp(-0.5\chi^2_{4,1-\alpha_2}).$$

The global level of the test is given by

$$\Pr_{H_0}(p_1 \leq \alpha_1) + \Pr_{H_0}(C(p_1,p_2) \leq c_{\alpha_2},\ \alpha_1 < p_1 \leq \alpha_0) = \alpha,$$

which turns out to be

$$\alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^1 I_{[C(x,y) \leq c_{\alpha 2}]} dxdy = \alpha$$

when p_1 and p_2 are independent and uniformly distributed.

Last, overall *p*-values have been derived for recursive combination tests, but there are several definitions and not all may be easy to compute in practice (19, 42).

Weighted sum of normal deviates

If the test statistics used at each stage are independent standard normal deviates Z_i, Lehmacher and Wassmer have proposed a combination function of the form (7)

$$Z = w_1 Z_1 + w_2 Z_2,$$

where weights $w_i \in (0, 1)$ verify $w_1^2 + w_2^2 = 1$. For a two-group balanced trial in every stage and with known equal variances, weights $w_i = \sqrt{\frac{n_i}{n_1 + n_2}}$ are optimal in the sense that, without sample size modification, the combination test is equal to the uniformly most powerful test.

As the combination rule has to be fixed in advance, one may use the original weights w_i for combining stage-wise Z_i despite sample size modification, and this has been shown to control for the overall type I error rate α (7). In such a case, the weights, however, are suboptimal and may lead to a loss of power as compared to a fixed sample test with the actual (and not the original) sample size. Bauer and Köhne also reported a small loss of power calculations with Fisher's product criterion, as compared to the uniformly most powerful test (4).

Remarking that $Z_i = \Phi^{-1}(1 - p_i)$, where $\Phi(.)$ denotes the cumulative distribution function of the standard normal distribution, the combination test described here can be expressed in terms of combination of *p*-values as:

$$C(p_1,p_2) = 1 - \Phi[w_1 \Phi^{-1}(1 - p_1) + w_2 \Phi^{-1}(1 - p_2)].$$

The final critical boundary of the test is computed to ensure

$$\Pr_{H_0}(Z_1 \geq z_{\alpha_1}) + \Pr_{H_0}(z_{a_0} \leq Z_1 < z_{\alpha_1},\ Z \geq z_{a_2}) = \alpha,$$

z_α being the $(1 - \alpha)$-quantile of the standard normal distribution.

Conditional error function

Another concept has been used to design flexible trials (5). Let us consider that the test statistic obtained at stage 1 is Z_1, with probability density function $\Phi(z_1)$. The conditional error function is any pre-specified increasing (if larger values of Z_1 indicate evidence against H_0) function $A(z)$ with range [0, 1], and satisfying

$$\int_{-\infty}^{+\infty} A(z)\Phi(z)dz = \alpha.$$

At the second stage, any test statistic independent of Z_1 and performed at the level $A(z_1)$ will allow control of the overall type I error rate. The conditional error function thus represents the probability of a type I error at the final stage, given observation of $Z_1 = z_1$ at the interim analysis.

Several conditional error functions have been considered in the literature, and little is known about optimal choice of the conditional error function (5, 40). For two-stage designs, Fisher's combination test has been compared to the circular conditional error function presented by Proschan and Hunsberger (5), and power and sample sizes were found to be almost identical (43). The combination tests described in previous sections may also be expressed in terms of conditional error functions. For instance, the one presented in previous section corresponds to the so-called linear conditional error function:

$$A(z_1) = \begin{cases} 0 & \text{if } z_1 < z_{\alpha_0} \\ 1 - \Phi\left(\dfrac{z_{\alpha_2}\sqrt{n_1 + n_2} - z_1\sqrt{n_1}}{\sqrt{n_2}}\right) & \text{if } z_0 \leq z_1 < z_{\alpha_1} \\ 1 & \text{if } z_1 > z_{\alpha_1} \end{cases}$$

Estimation

The methodology of flexible designs rests on control of the type I error rate while allowing design adaptations. Methods have thus focused more on statistical testing than on estimation. While the literature of flexible designs has increased in recent years and become highly developed, estimation remains a challenging issue, both from a theoretical and a practical point of view. Currently, several estimators are available, which are mostly biased. The experimenter is left with the choice of one of these estimators, whereas biased estimation may be difficult to interpret. For reviews on estimation in flexible designs, see (44, 45).

Point estimates

When sample size is re-evaluated at the interim analysis, the maximum likelihood estimator is typically mean biased. The bias depends on the alternative, the stopping rule, and the adaptation rule, and is therefore practically unknown. The possibility of early stopping also produces mean bias estimates, with positive or negative bias according to the situation. For instance, stopping for early rejection typically produces positively mean biased estimates, while stopping for futility produces negatively biased estimates. A limit for the mean bias has been derived, which is 40% of the standard error of the first-stage estimate; this limit may be improved if conditions are added on minimum and maximum sample sizes (46).

Some mean unbiased estimators have also been considered, but they use only part of the data and can show very large mean square errors. More reasonable mean square errors have been obtained with median unbiased estimators, that have smaller bias than the maximum likelihood estimator and comparable mean square errors. Both perform similarly well in terms of mean square error, as long as sample size adaptations are not too extreme (44).

Confidence intervals

Numerous controversial definitions of confidence intervals for flexible designs are available, and these can differ considerably depending on the possibility of early stopping or not. Among others, repeated confidence intervals and monotone confidence intervals have been proposed (19). Confidence intervals can be based on the weighted z-score test statisctics or on the likelihood ratio test statistics, the former being centred on the median unbiased estimate. All methods may be adapted to ensure that the flexible confidence interval includes the maximum likelihood estimate, or even the whole naive confidence interval (i.e. not adjusted to account for the sequential adaptive procedure). To date, there have beenvery few works aimed at comparing the different methods, but more detailed discussion can be found in the references cited above (44, 45).

Paths for future research

Flexible designs are an extension of group sequential designs. By flexibility, we mean the possibility of design adaptations using all possible sources of information, at the time of planned or unplanned interim analyses. In general, flexible designs have been shown to be inefficient as compared to group sequential designs (47). This is, however, the price to be paid for having the possibility of re-planning the ongoing trial. The need for adaptivity thus has to be crucial if one is to opt for a flexible design and not a traditional group sequential trial. This may be the case when using internal pilot studies, when nuisance parameters are particularly difficult to anticipate, or when several drug development phases are combined, as in a phase II/III seamless trial.

Advantages of adaptive designs are recognised for early-phase trials, but more debated for phase III trials (2). Pro-flexible trials insist on the maximum reductions that can be achieved, in terms of exposed patients and total costs of drug development, while shortening the time to decision making. For instance if an interim analysis shows that the probability of final rejection is low because the variance was underestimated at the trial planning stage, it may be worth deciding to increase the sample size instead of continuing the present trial, and then setting up a new one with the right variance. Flexible designs make it possible to allocate resources on the basis of early data, and therefore to reduce the risks of the entire clinical programme. If the number of patients does not necessarily have to be lower as compared to a group sequential design, combined phase II/III trials

usually require fewer patients than separate trials. It has also been underlined that adaptive designs can allow the targeting of subsets of patients who may benefit from the experimental treatment, using, for example, genomic or proteomic markers. At least the flexible trial methodology provides a safe theoretical framework (in terms of control of the type I error rate) for common practice protocol amendments.

On the contrary, detractors of adaptive designs point out their vulnerability to unblinding and bias, even though this also applies to a lesser extent to group sequential designs. In fact, all limitations of group sequential trials may also apply to flexible designs. It has also been argued that flexible designs over-emphasise statistical significance over clinical significance. Combination tests can lead to absurd results because of different weighting of equally informative observations (48). Examples may be found where the hypothesis $\mu \leq 0$ is rejected even when the average of observations is negative. There are, however, methods to overcome this problem, such as the dual test, where the null hypothesis is rejected if both the combination test and the naive test are significant at the pre-specified level (14). From a practical point of view, flexible designs complicate the trial's logistics, and the need for decisions on whether or not to adapt the design may ultimately increase the role of data and safety monitoring boards (DSMBs). It is not certain that these boards are ready to assume such increased responsibility, or that trial sponsors are ready to transfer the responsibility for taking important decisions. One solution could be to have representatives of the sponsor, who are not implicated in the particular trial, in the DSMBs, but this also needs further discussion.

While the methodology is already quite developed, adaptive designs are still a young and challenging area of clinical trials. From a statistical point of view, we have already underlined that estimation remains quite controversial, and solutions are still awaited. Other issues concern survival data analysis. Flexible designs are often seen as better suited for trials with immediately available endpoints. However, many trials are conducted over very long time periods in fields like cancer, with time-to-event endpoints. These trials are likely to benefit from the possibility of design adaptation, because misspecification of nuisance parameters would have dramatic consequences, because knowledge about the disease is more likely to be modified over long periods than over short periods of time, and because these fields are also the ones in which genomics and proteomics research, to find targeted therapies, is most advanced (21). If flexible designs can be used for survival data exploiting the independent increments property of log-rank statistics (21, 49), it has been argued that the use of information on covariables or surrogate endpoints of patients who did not fail at the interim analysis is not permissible since it might not be independent of the failure times observed at the second stage (50). This stresses the question of flexible designs with dependent data, which may also arise in phase II/III seamless trials. Recent work has been devoted to this issue, and solutions may be found, albeit at the cost of less simple calculations (42). This remains, nevertheless, one of the most challenging aspects of flexible designs.

From a more practical point of view, many methods are available, and it is not easy to know which one to choose. Although a certain number of adaptive or flexible designs have been successfully conducted and published in the medical literature (51), we believe that the more applications are published, the simpler it will become to plan new adaptive trials. In particular, only experience can help the community to set up practical guidelines. Moreover, this would help disseminate the methodology and show that there are situations in which adaptive designs are useful. Using adaptive designs also leads to changes in the way statisticians and clinical teams interact for trial planning. The role of simulations may be increased at that stage, to study different adaptation scenarios for example. In general, using adaptive designs increases the need for efficient infrastructures for data monitoring and reviewing.

Adaptive designs are most efficient as part of an adaptive drug development strategy rather than as simple isolated flexible trials. When using an adaptive design, it is fundamental to preserve the credibility of the trial and adaptations should be designed thoroughly. In any case, adaptive designs are not a convenient means of rescuing poorly planned trials. On the contrary, planning and conducting a flexible trial is more demanding, and may imply new tasks for all the teams involved.

References

1. Galo P, Chuang-Stein C, Draglin V, Gaydos B, Krams M, Pinheiro J. Adaptive designs in clinical drug development – an executive summary of the PhRMA working group. Journal of Biopharmaceutical Statistics 2006; 16: 275-283.
2. Kuehn BM. Industry, FDA warm to "adaptive" trials. JAMA 2006; 296: 1955-1957.
3. Bauer P. Multistage testing with adaptive designs (with discussion). Biometrie und Informatik in Medizin und Biologie 1989; 20: 130-148.
4. Bauer P, Köhne K. Evaluation of experiments with adaptive interim analyses. Biometrics 1994; 50: 1029-1041.
5. Proschan MA, Hunsberger SA. Designed extension of studies based on conditonal power. Biometrics 1995; 51: 1315-1324.
6. Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. Biometrics 1999; 55: 853-857.
7. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. Biometrics 1999; 55: 1286-1290.
8. FDA. Innovation or stagnation? - Challenges and Opportunity on the Critical Path to New Medical Products 2004. URL http://www.fda.gov/oc/intiatives/criticalpath/whitepaper.html.
9. FDA. Critical Path Opportunity List 2006. URL http://www.fda.gov/oc/initiatives/criticalpath/.
10. O'Neil RT. FDA's critical path initiative: a perspective on contributions of biostatistics. Biometrical Journal 2006; 48: 559-564.
11. Gottlieb S. Speech before the 2006 Conference on Adaptive Trial Design, Washington, DC 2006. URL http://www.fda.gov/oc.speeches/2006/trialdesign0710.html.
12. Rosenberg WF. Randomized play-the-winner clinical trials: review and recommendations. Controlled Clinical Trials 1999; 20: 328-342.
13. Bery DA. Bayesian statistics and the efficacy and ethics of clinical trials. Statistical Science 2004; 19: 175-187.
14. Posch M, Bauer P, Brannath W. Issues in designing flexible trials. Statistics in Medicine 2003; 22: 953-969.
15. Bauer P, Brannath W. The advantages and disadvantages of adaptive designs for clinical trials. Drug Discovery Today 2004; 9: 351-357.
16. Müller HH, Schäfer H. A general statistical principle for changing a design any time during the course of a trial. Statistics in Medicine 2004; 23: 2497-2508.
17. Bauer P, Brannath W, Posch M. Flexible two-stage designs: an overview. Methods of Information in Medicine 2001; 40: 117-121.
18. Hedges LV, Olkin I. Statistical Methods for Meta-analysis. NewYork; Academic Press 1985.
19. Brannath W, Posch M, Bauer P. Recursive combination tests. Journal of the American Statistical Association 2002; 97: 236-244.
20. Posch M, Bauer P. Interim analysis and sample size reassessment. Biometrics 2000; 56: 1170-1176.
21. Desseaux K, Porcher R. Flexible two-stage design with sample size reassessment for survival trials. Statistics in Medicine 2007 (in press).
22. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. Statistics in Medicine 1990; 9: 65-71.
23. Kieser M, Friede T. Re-calculating the sample size in internal pilot study designs with control of the type I error rate. Statistics in Medicine 2000; 19: 901-911.
24. Gould AL. Sample size re-estimation: recent developments and practical considerations. Statistics in Medicine 2001; 20: 2625-2643.
25. Jennison C, Turnbull BW. Mid-course sample size modification in clinical trials based on the observed treatment effect. Statistics in Medicine 2003; 22: 971-993.
26. Andersen PK. Conditional power calculations as an aid in the decision whether to continue a clinical trial. Controlled Clinical Trials 1987; 8: 67-74.
27. Shen Y, Fisher L. Statistical inference for self-designing clinical trials with a one-sided hypothesis. Biometrics 1999; 55: 190-197.
28. Denne JS. Sample size recalculation using conditonal power. Statistics in Medicine 2001; 20: 2645-2600.
29. Müller HH, Schäfer H. Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. Biometrics 2001; 57: 886-891.
30. Bauer P, Röhmel J. An adptive method for establishing a dose-response relationship. Statistics in Medicine 1995; 14: 1595-1607.
31. Bauer P, Röhmel J, Maure W, Hothorn L. Testing strategies in multi-dose experiments including active control. Statistics in Medicine 1988; 17: 2133-2146.
32. Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. Statistics in Medicine 1999; 18: 1833-1848.
33. Lawrence J. Strategies for changing the test statistics during a clinical trial. Journal of Biopharmaceutical Statistics 2002; 12: 193-205.
34. Kieser M, Schneider B, Friede T. A bootstrap procedure for adaptive selection of the test statistic in flexible two-stage designs. Biometrical Journal 2002; 44: 641-652.
35. Kieser M, Bauer P, Lehmacher W. Inference on multiple endpoints in clinical trials with adaptive interim analyses. Biometrical Journal 1999; 41: 261-277.
36. Hommel G. Adaptive modifications of hypotheses after an interim analysis. Biometrical Journal 2001; 43: 581-589.
37. Liu Q, Pledger GW. Phase 2 and 3 combination designs to accelerate drug development. Journal of the American Statistical Association 2005; 100: 493-502.
38. Bretz F, Schmidli H, Konig F, Racine A, Maurer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: general concepts. Biometrical Journal 2006; 48: 623-634.

39. Schmidli H, Bretz F, Racine A, Maurer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: applications and practical considerations. Biometrical Journal 2006; 48: 635-643.
40. Posch M, Bauer P. Adaptive two stage designs and the conditional error function. Biometrical Journal 1999; 41: 689-696.
41. Chang M. Adaptive design method based on sum of p-values. Statistics in Medicine 2007; 26: 2772-2784.
42. Liu Q, Pledger GW. on design and inference for two-stage adaptive clinical trials with dependent data. Journal of Statistical Planning and Inference 2006; 136: 1962-1984.
43. Wassmer G. A comparison of two methods for adaptive interim analyses in clinical trials. Biometrics 1998; 54: 696-705.
44. Brannath W, König F, Bauer P. Estimation in flexible two stage designs. Statistics in Medicine 2006; 25: 3366-3381.
45. Chang M. Adaptive Designs: Theory and Implementation Using SAS and R. Boca Raton; Chapman & Hall/CRC 2007.
46. Liu Q, Proschan MA, W PG. A unified theory of two-stage adaptive designs. Journal of the American Statistical Association 2002; 97: 1034-1041.
47. Tsiatis AA, Metha C. On the efficiency of the adaptive design for monitoring clinical trials. Biometrika 2003; 90: 367-378.
48. Burman CF, Sonesson C. Are flexible designs sound? Biometrics 2006; 62: 664-683.
49. Schäfer H, Müller HH. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. Statistics in Medicine 2001; 20: 3741-3751.
50. Bauer P, Posch M. (Letter to the editor in: Schäfer H, Müller H-H, Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections, Statistics in Medicine 2001; 20: 3741-3751). Statistics in Medicine 2004; 23: 1333-1334; authors' reply 1334-1335.
51. Bauer P, Einfalt J. Application of adaptive designs – a review. Biometrical Journal 2006; 48: 493-506.