

# **Guideline on Adaptive Designs for Clinical Trials**

**(Draft for Public Review)**



**Center for Drug Evaluation, NMPA**

**May, 2020**

## Table of Contents

<b>1. Introduction .....</b>	<b>1</b>
<b>2. Principles for Adaptive Designs.....</b>	<b>3</b>
<b>2.1 Validity.....</b>	<b>3</b>
<b>2.2 Integrity .....</b>	<b>4</b>
<b>2.3 Feasibility.....</b>	<b>6</b>
<b>3. Commonly Used Adaptive Designs .....</b>	<b>7</b>
<b>3.1 Group Sequential Design.....</b>	<b>8</b>
<b>3.2 Sample Size Re-Estimation .....</b>	<b>9</b>
<b>3.3 Two-stage Seamless Adaptive Design.....</b>	<b>11</b>
<b>3.4 Adaptive Enrichment Design .....</b>	<b>13</b>
<b>3.5 Master Protocol With Adaptive Design .....</b>	<b>15</b>
<b>3.6 Multiple Adaptive Design.....</b>	<b>17</b>
<b>4. Special Considerations for Adaptive Design.....</b>	<b>18</b>
<b>4.1 Bayesian Adaptive Designs.....</b>	<b>18</b>
<b>4.2 Simulation-based Adaptive Designs .....</b>	<b>21</b>
<b>5. Regulatory Considerations .....</b>	<b>23</b>
<b>5.1 Communication With Regulatory Agency.....</b>	<b>23</b>
<b>5.2 Documentation Requirements .....</b>	<b>23</b>
<b>5.3 Other Considerations .....</b>	<b>24</b>
<b>Reference .....</b>	<b>26</b>
<b>Appendix 1 Glossary .....</b>	<b>28</b>
<b>Appendix 2 Examples of Adaptive Designs .....</b>	<b>31</b>

# **Guideline on Adaptive Designs for Clinical Trials**

## **1. Introduction**

This guideline focuses on the important concepts and principles for adaptive designs in clinical trials. It provides guidance to sponsors and applicants submitting investigational new drug applications, new drug applications, biologics licensing applications, or supplemental applications on whether and/or how to appropriately use adaptive designs for clinical trials. Some commonly used adaptive designs will be discussed from a regulatory perspective, including recommendations for the sponsor to consider and some specific requirements. For adaptive designs submitted by the sponsor, the regulatory authorities will provide specific advice after a comprehensive assessment based on the submitted materials and various factors specific to the clinical trial.

As adaptive designs may involve a variety of statistical approaches, the sponsor should also refer to other relevant ICH guidelines and domestic guidelines to facilitate the design of adaptive clinical trials.

A fixed sample design, commonly referred to as a traditional design, is widely used in confirmatory trials. In a fixed sample design, the efficacy analysis is only performed at the end of the trial, with no analyses or design modifications conducted during the study. However, many confirmatory trials are designed on the basis of limited historical data which may result in large uncertainties. Modifying the trial based on data accumulated during the trial to adjust for deviations from the original design assumptions has become an important approach. An adaptive design makes certain modifications to the fixed sample trial. In this guideline, an adaptive design is defined as a clinical trial design that allows for prospectively planned modifications at an interim analysis based on accumulating data from subjects in the

trial. This modification is also called adaptive modification. The plan for these adaptive modifications must be prespecified in the trial protocol and statistical analysis plan before the start of the clinical trial.

The group sequential design may be considered one of the first adaptive designs applied to clinical trials. More recently, adaptive designs for sample size re-estimation have become increasingly used. After the EU EMA in 2007 and the US FDA in 2010 each issued a regulatory guidance for adaptive designs, adaptive designs were gradually promoted and developed into more types, from dose selection to more complex designs involving multiple target populations, multiple hypotheses, and multiple endpoints. With the growing new theoretical methods and increasing application experiences, more and more adaptive designs have been applied in clinical trials, covering almost all stages of drug development and a wide array of disease areas.

Given that many clinical trials fail due to limited prior information at the design stage, adaptive designs can greatly increase trial success rates by modifying the design based on data accumulated during the trial to improve upon the initial design. An adaptive design can also improve the efficiency of the trial, such as reducing the sample size, shortening the time interval between different study phases, selecting more appropriate endpoints and target populations, and more efficient use of information on each subject. In addition, complex adaptive designs can handle multiple trial objectives, multiple investigational agents, and multiple diseases simultaneously within a single trial.

Although adaptive designs have the potential advantages mentioned above, due to its complexity, it also brings many challenges and problems to the trial design, data analysis and interpretation of results, as well as, implementation of the trial. These challenges and problems include control of the overall type I error rate, choice

of an analysis model, estimation of the treatment effect, as well as the potential operational bias created by complications in trial implementation, all of which may limit the successful application of adaptive designs.

This guideline primarily focuses on the use of adaptive designs in confirmatory clinical trials of chemical drugs, biological products and traditional Chinese medicines, although the concepts are also relevant for exploratory studies.

## **2. Principles for Adaptive Designs**

Before deciding whether to use an adaptive design, the advantages and disadvantages compared to a traditional design should be comprehensively evaluated. Specific considerations include the complexities of an adaptive design in terms of design, conduct and statistical analysis, as well as the resulting non-avoidable operational bias and other various challenges that may be introduced in the trial conduct. Whether to adopt an adaptive design requires comprehensive consideration of many factors, particularly those affecting its validity, integrity and feasibility.

### **2.1 Validity**

The validity of adaptive designs refers to utilizing an appropriate statistical analysis method that will not cause bias in the estimation of efficacy. The validity of a trial is about the credibility, interpretability and persuasiveness of its results. Maintaining the validity of the trial means there should be correct statistical inference methods applied; e.g. how to calculate the adjusted p values, how to estimate effect sizes and confidence intervals, and how to measure the consistency of treatment effects at different stages.

Since this guideline focuses on confirmatory trials that support registration, the overall type I error rate of the trial is required to be strictly controlled at a two-

sided 0.05 (or one-sided 0.025) level. The most important criterion for judging whether an adaptive design is reasonable is whether the statistical methods used can control the overall type I error rate. For some adaptive designs, such as using two-sided tests, since the p values at different stages cannot reflect the direction of the comparisons between groups, it may make the final overall p values difficult to interpret. To avoid this, a one-sided test may be used instead. However, for other adaptive designs, such as asymmetric two-sided assumption, a two-sided test would be the more appropriate choice. Adaptive modifications should adjust the type I error rate of the trial, except for in some special cases.

Adaptive designs may involve multiple target populations, multiple hypotheses, multiple endpoints, or multiple tests at the same time, such that there are more rigorous requirements for the validity of statistical analyses. If there is no corresponding valid and effective statistical method for adaptive modifications, the design should not be adopted. Furthermore, due to the complexity of the adaptive design, there may be no applicable theoretical formulation for statistical inference in some cases, and the validity of the statistical approach needs to be verified based on simulation methods to some extent, which may increase additional uncertainty in the design.

The analysis of adaptive designs needs to pool data from multiple stages. The inconsistency of efficacy estimates across stages will not only make statistical inference based on the pooled data difficult, but may also make the trial results difficult to interpret. In addition, many adaptive modifications are aimed at achieving statistically positive results, however, if the final statistical test result is positive but the clinical benefit is too small, then it is not enough to support efficacy of the drug.

## **2.2 Integrity**

The integrity of adaptive designs refers to well control of the potential bias introduced by the trial operation. Maintaining the integrity of the trial means the adaptive modifications to be made according to a prespecified plan, and keeping the interim analysis results blinded to minimize operational bias.

Avoiding introduction of operational bias is the most essential requirement for all clinical trials. Since an adaptive design may involve modifications in many aspects of a clinical trial, the conduct of the remainder of the trial may be impacted, which increases the difficulty of maintaining trial integrity. Thus, all the interim analyses in an adaptive design trial should be conducted by an adaptation committee/ third-party experts and a statistical support team, both of which are independent of the sponsor, in order to ensure that the interim analysis results are not known to the sponsor, investigators, and subjects, so as not to affect the conduct of the remainder of the trial and introduce operational bias. In most cases, if the adaptive design is not particularly complex, the independent data monitoring committee (IDMC) can be responsible for the operation and implementation of the adaptive modifications; if the adaptive design is very complex, the sponsor may consider selecting experts with expertise in adaptive modifications to be on the IDMC; if the adaptive design is extremely complex and the data monitoring committee does not have sufficient relevant experience, it is necessary to set up an independent adaptive design committee. Depending on the degree of complexity, multiple statisticians could be included in the adaptation committee. Since adaptive modifications involve multiple factors, the most important task in implementation is setting up an effective firewall to prevent leakage of interim analysis results which may cause operational bias. To this end, the scheme of an adaptive design should include a complete operation process, especially on how to set access authority to relevant information. At the same time, in order to avoid the influence of uncontrolled factors on the trial results,

it is also necessary to consider how to avoid indirect inference of the interim analysis results based on the trial modifications. It should be noted that the adaptation committee should not disclose specific interim analysis results in their recommendations on trial modifications to the sponsor. The sponsor should prepare all required standard operating procedures, and incorporate all relevant procedures related to the adaptive modifications, and record all actual operation processes for review at the time of new drug application submission. All the above factors should be carefully considered during the design stage of the trial and strictly executed during the trial so as not to affect trial integrity and jeopardize the reliability of the trial conclusions.

### **2.3 Feasibility**

The feasibility of an adaptive trial is about whether the adaptive modification of the trial can be implemented in practice. Since adaptive designs are more complex than traditional designs and more difficult to implement and analyze, the following factors need to be considered before planning an adaptive design: the adaptive modification strategies should be able to ensure the validity and integrity of the trial; relative to the trial period, there should be sufficient time for the adaptive modifications and conduct of the remainder of the trial based on the analysis results of the cumulative data; interim data collection and data cleaning should be able to be completed quickly, so as to complete the interim analysis according to the scheduled plan without needing to suspend subject recruitment; should be able to quickly modify the randomization procedures/drug supply systems, should have adequate drug supply management capabilities to afford increased drug supplies, the data capture systems for adaptive designs should be prepared in advance; smooth and effective communication with relevant parties should be ensured; validated software should be available to complete complex designs and calculations of



relevant analyses, to meet the needs of adaptive modifications and implementation of the trial operations. At the same time, during the design stage of trial, the sponsor can communicate with the investigators and propose a target list of the trial based on clinical considerations, to assess the feasibility of the considered adaptive design in practice. If relevant adaptive modifications are difficult to implement, other designs should be considered.

In summary, if an adaptive design is planned to be adopted, it needs to be carefully evaluated with respect to its advantages. If a decision cannot be made, simulation methods could be used to compare with traditional designs to evaluate design efficiencies of adaptive designs, and select the better design. If the advantages of an adaptive design are limited after evaluation, their usages should be carefully considered.

### **3. Commonly Used Adaptive Designs**

An adaptive design is a clinical trial design that allows for prospectively planned modifications to the design based on accumulating data from subjects in the trial under the premise of ensuring validity and integrity. On the one hand, adaptive modifications are made "according to a prespecified plan" rather than ad-hoc modifications; on the other hand, adaptive modifications are a self-learning process, that is, through the continuous learning from cumulative data, the trial protocol is modified correspondingly to accommodate the evolving situation. Thus, an adaptive design aims to improve the ongoing clinical trial rather than waiting until the end of the trial to realize potential deficiencies in the design that may have led to its failure.

Adaptive designs involve a wide range of applications. Due to the limited scope, this guideline will only discuss several commonly used adaptive designs, including group sequential design, sample size re-estimation, two-stage seamless

adaptive design, adaptive enrichment design, adaptive master protocol trial design, and multiple adaptive design. The principles and methods of these designs are also applicable to most other adaptive designs. In addition, the methods discussed in this section will be illustrated in several hypothetical cases (see Appendix 2).

### **3.1 Group Sequential Design**

The group sequential designs allow for one or more prospectively planned interim analyses during the trial with prespecified decision-making criteria. Generally there are four types of decisions: ① Stop the trial for efficacy; ② Stop the trial for futility; ③ Stop the trial due to safety concerns; ④ Continue the trial. The timing of the interim analysis can be based on calendar time or on the proportion of cumulative data, such as the proportion of subjects enrolled or the proportion of events occurred. If the interim analysis is planned for efficacy assessment which may potentially lead to stopping the trial early for futility or superiority, the Type I error rate should be adjusted for each analysis such that the overall Type I error rate is controlled at the 2-sided 0.05 (or 1-sided 0.025) level. Common methods to control the Type I error rate include the Pocock method, O'Brien & Fleming method, and Lan & DeMets method. Since only some of the data are used in the interim analysis, the results may still have large uncertainties. Thus, more conservative methods for efficacy boundaries should be considered to increase the reliability of conclusions. The futility rules in the group sequential design may be binding or non-binding. Binding boundaries may decrease the probability of rejecting the null hypothesis. Therefore, the superiority boundary can be appropriately relaxed to increase the probability of a positive result while controlling for overall Type I error. The trial must be terminated once the futility binding boundary is crossed at the interim analysis. For non-binding boundaries, the Independent Data Monitoring Committee

can recommend the trial to continue based on broader considerations, even when the results of the trial cross this boundary.

The timing of the interim analysis should also be carefully considered. If there is a possibility to stop the trial for efficacy early in the group sequential design, the choice of timing should consider whether the interim data are sufficient to provide reliable evaluation of efficacy and safety including key secondary endpoints and key subgroups analyses. If the interim analysis is to assess the safety and futility of the drug, the timing consideration should focus on how to maximize protection of the subjects.

### **3.2 Sample Size Re-Estimation**

Sample size re-estimation refers to recalculation of sample size according to a prespecified rule based on accumulating trial data at an interim analysis to ensure that the final statistical test could achieve the prespecified criteria or modified criteria with the overall Type I error rate controlled.

The initial sample size calculation is usually based on factors such as effect size, variability of the primary endpoint, duration of trial follow-up, and dropout rate, which often come from historical data. In most cases, the information available for sample size estimation at the design stage is often insufficient and may lead to inaccurate estimation of the sample size. Thus, sample size re-estimation in an adaptive design provides a potential approach to mitigate these issues.

Sample size re-estimation includes blinded sample size re-estimation and unblinded sample size re-estimation.

Blinded sample size re-estimation, also known as non-comparative analysis, refers to interim analyses that do not use information on the actual treatment groups, or do not conduct any analyses involving comparisons between groups, although

information on the treatment group is used, such as a pooled analysis of the data from the two treatment groups at the interim analysis.

In blinded sample size re-estimation, the estimation of key parameters used in the sample size calculation (eg, pooled variance or standard deviation) are based on the accumulating data, which are then used to re-estimate the sample size. Since the interim analysis does not involve comparison of efficacy between groups, it is generally not necessary to adjust the type I error rate. This method is relatively easy to implement, and does not generally introduce any operational bias. In addition, the relevant statistical methods are relatively mature. It is important for blinded sample size re-estimation to be pre-planned during the trial design stage.

Unblinded sample size re-estimation, on the other hand, is also known as comparative analysis, in which treatment group information (including the actual name of each group or group code) is used in the interim analysis. Such an analysis involves the comparison between groups.

The unblinded re-estimation of sample size involves estimation of parameters key to the sample size calculation (eg, effect size per group) based on the cumulative data and treatment group information, which are then used to re-estimate the sample size. Since the interim analysis involves comparison of the efficacy data between treatment groups, it is usually necessary to adjust the type I error rate accordingly.

Unblinded sample size re-estimation should be pre-specified in the protocol, including the timing of re-estimation, what the decision-making criteria are, what method to use for the re-estimation, and how to adjust  $\alpha$  such that the overall Type I error rate is controlled. Additionally, it should be pre-planned who will conduct the unblinded analysis, and finally who will perform the entire procedure. Special attention should be paid to the fact that a sample size re-estimation is generally recommended only once in a trial. When the re-estimated sample size is less than

the sample size of the initial design, adjustment for sample size reduction is usually not accepted unless there is a special justification.

An adaptive design with an unblinded sample size re-estimation requires multiple considerations. For example, is it necessary to conduct unblinded sample size re-estimation in the setting where reliable historical data is available? Which strategy has more advantages when considering benefit-cost of unblinded sample size re-estimation against the increase in sample size from the original design (ie, due to adjustment of type I error rate)? Can the interim analysis be completed in a timely manner? Is it possible that there is insufficient time to adjust the trial due to the high enrollment rate? At what point to conduct the interim analysis? Therefore, the design should be based on the trial characteristics. It is recommended to choose the most appropriate method after careful consideration of these factors and more.

There are many resources in the literature available for sample size re-estimation methods such that a suitable method can be selected for each specific case.

### **3.3 Two-stage Seamless Adaptive Design**

The two-stage seamless adaptive design refers to the division of a trial into two stages. An interim analysis is performed at the end of stage 1, after which the Stage 2 trial may be adaptively modified based on predefined criteria. Seamless design is usually divided into operationally seamless design and inferentially seamless design. Operationally seamless design excludes first-stage trial subjects from the main analysis, thus with no need to adjust for the overall Type I error rate. Inferentially seamless design needs to include data from all subjects enrolled in both stages of the trial, with appropriate adjustment to control the overall Type I error rate.

According to the trial objective(s) and endpoint(s), there are generally four types of two-stage designs: same objective(s)/same endpoint(s), same

objective(s)/different endpoint(s), different objective(s)/same endpoint(s) and different objective(s)/different endpoint(s). Any two-stage seamless adaptive design can be appropriately classified into one of these four types, and the appropriate type can be selected for the specific trial.

If the number of treatment arms is the same in both stages, the group sequential design can be considered a special case of the same objective(s)/same endpoint(s) seamless design. In two-stage seamless trials, it is common to have Phase I/II and Phase II/III seamless adaptive designs. The former one is commonly used in exploratory trials with a biomarker explored in Stage 1 and early efficacy signal explored in Stage 2. The latter one is commonly used in confirmatory trials with dose selection performed in Stage 1 and efficacy confirmed in Stage 2.

Independent phase II trials typically include multiple trial arms, such as multiple dose levels for the same drug to select the appropriate dose level and decide whether to proceed to Phase III trials. Phase III trials are independent of Phase II trials, and the data from Phase II are not included in the analysis of Phase III. This approach does not adequately utilize data from the Phase II trial. Inferentially seamless adaptive designs include data from all subjects enrolled in both phases of the trial at the final analysis. It has many potential advantages, such as shortening the time interval between the end of Phase II and the start of Phase III, reducing the total sample size of the trial, shortening the duration of the trial, reducing the cost of the trial, and increasing the sample size for the final analysis. Due to the longer follow-up period from subjects enrolled in stage 1, it may provide an earlier readout for long-term safety of the drug.

Multiple factors need to be considered in a seamless phase II/III adaptive design. Given that stage 1 results may not be comprehensive, dose selection should involve many factors. There may be difficulties in the design, operation as well as

implementation arising from the adaptive design; it is generally not appropriate to have a two-stage seamless adaptive design if the investigational product is not well-understood. There are also situations where the use of a two-stage seamless adaptive design may pose a greater risk. For example, the primary endpoint of a phase III trial requires a long follow-up period. Dose selection in Stage 1 may only be based on surrogate endpoints. When the relationship between surrogate endpoints and primary endpoints is not high or in fact weak, the dose selected in phase II to use in phase III based on surrogate endpoints may bring great uncertainty. Also, some issues may arise in the scenario where the primary endpoint needs a longer duration of follow up but the enrollment period is short. Enrollment may need to be put on hold to wait for the results of the interim analysis in order to prevent too many subjects from entering the unselected dose group in Phase III.

The approach for the two-stage adaptive design discussed above can also be applied directly to other similar trials, such as Stage 1 involving the selection of different drugs, or the selection of combinations or single agents.

### **3.4 Adaptive Enrichment Design**

Under the framework of a two-stage seamless adaptive design, an adaptive enrichment design refers to the adaptation according to pre-defined criteria at the end of Stage 1 to determine the target population for Stage 2 based on the interim analysis results. Stage 2 of the trial may continue to enroll the overall population, or only enroll subpopulations after adaptive modification(s). The sample size for the overall population may also be increased, which naturally increases enrollment in the subpopulations. The final analysis may be based on either the overall population only, subpopulation only, or both the overall population and subpopulation, the importance of which is determined by how  $\alpha$  is allocated. The final analysis of the

trial will include all subjects enrolled in both phases with the overall Type I error rate controlled using an appropriate adjustment method.

If the drug is known to work only in a certain subpopulation, the clinical trial should recruit only subjects from that subpopulation. However, more common in practice is that there may be a greater effect in a particular subpopulation and it is unclear whether there is a meaningful effect in the overall population. In this case, if the investigational product has a large enough effect in the overall population, enrollment of the subpopulation alone will lose the opportunity to show a treatment effect in the overall population. If the investigational product has little effect in the overall population but is effective in a certain subpopulation, it is highly unlikely that subjects enrolled in the overall population will have the expected positive results. In this case, the opportunity to show efficacy in the subpopulation may be lost if the target population cannot be determined accordingly. The use of a two-stage seamless adaptive design to select the target population based on the accumulating data in the trial itself (ie, Stage 1) facilitates identification of the target population in a scientific manner and increases the success rate of drug development.

Because selection of the target population in an adaptive design involves the overall population and subpopulations, if the Stage 1 interim analysis uses an unblinded between-groups comparison, the statistical assumptions for the two populations and corresponding statistical methods should be clearly defined to control the overall Type I error rate.

Selection of the target population can be based on various criteria such as disease characteristics, prognostic biomarkers, or predictive biomarkers. In general, the design and implementation of the trial will become relatively simple if selection of the target population is based on established disease-related characteristics or prognostic/predictive biomarkers. Currently, while there is an increasing number of



studies selecting for a target population by employing predictive biomarkers, the clinical value of many predictive biomarkers is not yet clear. If the trial is to use a completely new predictive biomarker to select the target population, there must be a corresponding diagnostic method. The diagnostic method must have been approved by the regulatory authorities. If not, it may require simultaneous development. If the data ultimately does not support the in-vitro diagnostic developed (ie, failure to obtain approval from the device regulatory agency for the marketing application), it will directly lead to unreliable conclusions regarding the test drug in the trial. In addition, if the cutoff threshold corresponding to the in-vitro companion diagnostic has not yet been well-established to define the subpopulation, some of the early enrolled subjects who are used to determine the threshold should be excluded from the final analysis to avoid difficulties in interpreting the final results. To more comprehensively understand the prediction of the biomarker and fully assess the trial outcomes, it is generally encouraged to include the non-target population in the study as well.

In the absence of sufficient knowledge of drug effects in the subpopulations, it is difficult and risky to decide whether to use an adaptive design to select the target population. If the effect in the target subpopulation is unknown, positive results may not be obtained in the overall population, or even if a positive result is obtained in the overall population, there may be a lack of effect in the non-target subpopulation. Both situations may lead to ethical issues. On the other hand, if the drug works in the overall population and several subpopulations but only one of them has been selected in the trial, that would also result in insufficient use of effective drugs.

### **3.5 Master Protocol With Adaptive Design**

The master protocol trial design refers to a master protocol containing multiple sub-protocols. Different sub-protocols can simultaneously evaluate the effect of a

drug for multiple diseases, or simultaneously evaluate the effects of multiple drugs for one disease, or simultaneously evaluate the effects of multiple drugs for multiple diseases. Every sub-protocol can be a single-arm trial, or a randomized controlled trial. If sub-protocols are randomized controlled trials, each sub-protocol may share a single control group or have a control group of its own. The master protocol trial also refers to clinical trials marked by patient-specific characteristics (eg, disease, histological type, molecular markers). The master protocol trial has many advantages, such as to maximize enrollment opportunities for patients and provide the most appropriate test drug. Common master protocol designs include basket trial, umbrella trial, and platform trial.

Master protocol with adaptive design refers to a design that includes one or more adaptive modifications in the master protocol. It can flexibly implement multiple adaptations, such as adding one or more new sub-protocols, stopping early for one or more sub-protocols, re-estimation of sample size, adjustment of hypothesis testing, primary endpoint(s) and primary statistical methods, or different adaptive modifications for different sub-protocols.

The master protocol trial should institute a central committee or board, for example, independent safety assessment committee, independent data monitoring committee, and/or independent review board. A central randomization system, electronic data collection system, central laboratory, and use of centralized case report forms, informed consent form and clinical monitors, etc. are required for implementation.

A basket trial design is used to evaluate the clinical efficacy of a drug for treatment of different disease types with the same biological characteristics. The design contains multiple sub-protocols in a master protocol, usually each of which is a single-arm trial targeting one or more disease types. In the field of oncology,

sometimes it is difficult to run individual traditional trials for each tumor histology. The basket trial is able to meet the needs brought by the advancement of molecular biology classification using gene sequencing and genome-wide analysis.

An umbrella trial design is used to evaluate the clinical efficacy of multiple drugs targeting the same disease or biomarker type. The design contains multiple sub-protocols in a master protocol, each of which may be a single-arm or randomized controlled trial targeting one or more drugs. Umbrella trials are commonly used to select candidates for confirmatory studies and can also be used as confirmatory studies.

The platform trial design is used to evaluate the clinical efficacy of multiple treatments for a variety of diseases. This design contains multiple sub-protocols in a master protocol, each of which is a randomized controlled trial and generally shares the same control group. The platform trial is usually maintained for a long term and allows for new drugs to be added to or dropped from the platform at any time. Furthermore, the comparator(s) may change over time.

A master protocol trial has many advantages, however, due to its complexity in planning, implementation, the establishment of structured management board(s) and especially the statistical analysis, many challenges exist. The master protocol trial should be carefully planned after thorough, in-depth, and meticulous evaluation of the various issues that may be involved in various aspects of the trial.

### **3.6 Multiple Adaptive Design**

The multiple adaptive design refers to a trial design in which two or more adaptive modifications are applied in a trial. The adaptive design methods discussed above may all be used simultaneously within the same clinical trial. For example, a clinical trial that determines the dose at the end of Stage 1 will then require a sample size re-estimation before selecting the target population.

In principle, if a clinical trial design contains multiple adaptive modifications, multiple adaptive designs can be considered as long as they meet the requirements of validity, integrity, and feasibility. Due to the complexity of multiple adaptive designs, it is suggested to carefully consider whether it is indeed necessary to introduce the many adaptive modifications in one trial.

## **4. Special Considerations for Adaptive Design**

### **4.1 Bayesian Adaptive Designs**

A Bayesian adaptive design refers to a trial design that uses Bayesian methods and also contains adaptive modifications. Bayesian methods are a class of statistical methods that combine the information/data summarized by a distribution function (prior distribution) with the data obtained from the current trial according to the Bayesian principle, to obtain a new distribution function (posterior distribution) to summarize this information/data, and make statistical inference based on this posterior distribution function. The information/data from previous trials can be based on the drugs to be tested in the current trial, and can also be based on other relevant drugs.

In clinical trials, the primary task is to obtain an accurate and reliable estimate of drug efficacy. Sometimes, a prior distribution can be used to summarize the information/data from previous trials to obtain an initial estimate of drug efficacy. Due to insufficient information/data or other uncertainties in previous trials, it is not possible to obtain an accurate and reliable estimate of efficacy based solely on the trial itself. Thus, more data needs be collected within the current trial. Based on the newly collected data, the initial estimate of efficacy (prior distribution) is updated and a new estimate (posterior distribution) is obtained. Estimates of efficacy obtained from Bayesian methods can often be regarded as weighted averages of

information/data from previous trials and data from the current trial in a specific manner, i.e., if there is no data from the current trial, the estimate of efficacy will be entirely based on the information/data from previous trial(s); if there is data from the current trial, the estimate of efficacy will be a weighted average. The weight of the data from the current trial will increase as the amount of data increases, approaching 1.

The adaptive designs based on frequentist theory discussed above in this guideline are mostly applicable to Bayesian adaptive designs as well. Since Bayesian methods use previous or related information/data in statistical inference, it naturally has certain advantages in specific situations. The flexibility of Bayesian methods lies in fact that some statistical models can be used to borrow relevant data. There are many situations where it is difficult to conduct a clinical trial alone with an appropriate sample size. Thus, it may be necessary to use Bayesian methods to borrow relevant data to obtain more credible conclusions. For example, borrowing data from adult clinical trials in pediatric clinical trials; borrowing data from similar disease indications given the inability to enroll enough patients in rare diseases; borrowing data from adjacent regions given enrolled patients are not sufficient in a certain region; borrowing data from previous trials to reduce the number of patients in the control group of a non-inferiority clinical trial. Bayesian methods can provide quantitative analyses and interpretations for these borrowing approaches.

Despite the superiority of Bayesian methods in certain respects, the biggest problem is the uncertainty of the statistical inference of the results. Using the same information/data from previous trials and data from the current trial, Bayesian inference may lead to different conclusions if either different prior distributions are selected or different parameter values are used even if the same prior distribution is selected. In addition, there are no acknowledged Bayesian methods to select the

decision criteria for final statistical inference. Given these issues, Bayesian methods are more currently used for exploration of drug doses in Phase I clinical trials, selection of subsequent development strategies in Phase II clinical trials, interim futility analyses, some predictive analyses in Phase 3 clinical trials, and in many other analyses that are not intended for the purposes of registration.

Due to the complexity of adaptive designs and the limitations of statistical methods based on frequentist theory, although Bayesian methods have their shortcomings, the use of Bayesian methods may be a more appropriate option in some designs. If Bayesian methods are used, sufficient prior information/data, literature and studies are needed to support the validity of the statistical model(s) used, including the selected prior distribution and the values of each parameter. In addition, due to the uncertainty caused by Bayesian inference based on the choice of prior distributions and parameter values, a large number of simulation results are needed to illustrate the operating characteristics of designs across hypothetical scenarios that might occur in practice. In particular, it is necessary to show whether the decision criteria defined based on posterior probabilities in the trial are valid via simulations; for example, the overall type I error rate corresponding to the statistical methods based on frequentist theory can be used to evaluate the selected decision criteria. Furthermore, it is also necessary to consider the feasibility of using Bayesian methods in practice, such as how to interpret the meaning of various statistical models to investigators, the meaning of decision criteria defined based on posterior probabilities, the interpretation of estimates of drug efficacy, whether the randomization based on adaptive probabilities of unequal responses will bring additional safety risks to subjects, and whether the delay caused by updating adaptive probabilities will make the actual operation of enrollment overly difficult. Here the adaptive probabilities of responses refer to the updated proportions of

patients randomized in the future through probabilities based on efficacy calculated using data of enrolled patients in each trial arm.

In view of the many challenges of Bayesian methods, if Bayesian methods are used, careful consideration of the various issues discussed above and advanced research and planning are needed.

#### **4.2 Simulation-based Adaptive Designs**

Adaptive designs based on simulation methods refer to exploring the validity of statistical inference made in adaptive trials through use of simulation methods. In clinical trials, statistical tests lead to statistical inference based on certain distribution theory or approximate normal distribution theory under certain statistical assumptions. These conditions required by the distribution theories or approximate normal distribution theories are generally met in traditional clinical trials. In order to tailor to the needs of drug research and development, many novel and complex trials are constantly emerging; for example, the master protocol trial involves multiple target populations, multiple hypotheses, multiple endpoints, and/or multiple tests at the same time, which poses new challenges in deriving distribution theories of the statistical tests. In many extremely complex trials, the conditions based on distribution theories may no longer be satisfied and the basis needed for establishing statistical inference can only be obtained using simulation methods.

The greatest advantage of statistical trial simulations is that it provides a better understanding of the operating characteristics within a hypothetical clinical trial scenario. Specific to the clinical trial simulations, the choice of simulation models and parameters is important such that they are appropriate to describe the scenarios in which the trials may occur, and the overall type I error rate is controlled. Given the many resources in the literature on statistical simulations, the computational

details of simulation methods, computer languages, simulation software, and control of simulation error will be not discussed here.

If there is no clear basis for theoretical distributions, then it is theoretically impossible to prove that the overall type I error rate can be strongly controlled under the null hypothesis in clinical trials. The overall Type I error rate involves the entire null hypothesis space, i.e., assuming that the treatment and control groups have the same efficacy, which theoretically has infinite possibilities, such that no single simulation can exhaustively assess all the scenarios for verification. It is necessary to consider excluding some obviously unreasonable scenarios in the simulations, and limit the evaluation to more realistic scenarios depending on disease characteristics and/or historical data. In this way, the simulation results based on the reduced null hypothesis space will still be reliable from a statistical perspective. In addition, besides considering how to choose the main parameters in the simulations, it is also necessary to consider many other factors such as nuisance parameters, enrollment rate, dropout/censoring rate, follow-up time and simulation accuracy rate. After selecting these parameters, various modifications involved in the adaptive design, as well as multiple target populations, multiple endpoints and/or multiple tests that may be involved are added in order to show that the proposed statistical methods can still control the type I error rates after multiplicity adjustments in clinical trials.

In view of the uncertainty of statistical inference based on simulation methods, it is prudent to carefully consider various factors unless the adaptive design is quite necessary and indeed has greater advantages compared to traditional designs. An adaptive design based on simulation methods can be considered in situations where sufficient medical literature, previous data and/or other evidence can support its need, and reliable simulation methods and corresponding results show that the adaptive design indeed has great advantages.



## **5. Regulatory Considerations**

In view of the complexity of adaptive designs, the sponsor should communicate with the regulatory agency as early as possible during the trial design in order to allow sufficient time to improve the plan accordingly.

### **5.1 Communication With The Regulatory Agency**

For adaptive designs aimed at exploratory objectives, it is not necessary to communicate with the regulatory agency. However, communication with the regulatory agency is needed if the trial may affect the safety of many subjects, eg, a master protocol trial with a large number of subjects; or early development is aimed at exploratory objectives, which may later evolve to confirmatory studies. Usually, it is necessary to communicate with regulatory agency in advance for adaptive designs in confirmatory studies such that there is sufficient time to consider the suggestions, concerns, and/or opinions of regulatory agency in the early stage of design, especially for designs that are complex and/or utilize new methods. Any documented agreements with regulatory agency should be reflected in the amended protocol.

### **5.2 Documentation Requirements**

The documents to be submitted by the sponsor should contain all the theories, literatures and data used to support the use of the adaptive designs for review by the regulatory agency. Preparation of the documentation should focus on the pre-specified adjustment plan and comprehensive discussions of the clinical meaningfulness, validity, integrity and feasibility.

Medical meaningfulness is an important factor in judging whether the use of the adaptive design is appropriate. The documentation should contain sufficient evidence to support clinical meaningfulness of the trial results after appropriate

adjustments. For example, after one or more adjustments, interpretation of the trial results may become quite difficult, or the trial results may eventually reach a statistically positive result without clinical meaningfulness.

Validity applies mainly to the statistical methods, of which the most important criterion is whether the statistical methods used can control the overall type I error rate at a 2-sided 0.05 (or one-sided 0.025) level. The documentation should include the pre-specified adjustment plans, all adjustment procedures and details, and all references cited. If the adaptive design is extremely complex and there is no specific theoretical formulation, it may need to be illustrated by simulation methods. The sponsor needs to consider during the planning stage whether the simulation results can be independently verified by a third party.

Integrity pertains to the operation and conduct of the trial, of which the criterion is that the design used will not introduce bias due to the trial operation or conduct. The documentation should include all operational procedures, especially how to set up a firewall to ensure that the analysis results will not be disclosed. Other relevant guidelines may be referred to for adaptive modifications that the data monitoring committee are in charge of.

Feasibility is aimed at assessing whether the potential adaptive modifications planned can be implemented in practice, which requires the sponsor to make a comprehensive evaluation.

The above are only the basic contents that should be included in the documentation. If the sponsor thinks that there are other materials that would facilitate communication with the regulatory agency, those may also be submitted.

### **5.3 Other Considerations**

In principle, the plans for adaptive modifications in an adaptive design must be prespecified within the trial protocol and statistical analysis plan before the

clinical trial starts. In general, non-prespecified modifications to a trial are not recommended. However, in the practice of clinical trials, it is sometimes necessary to make ad-hoc modifications based on the data. After careful consideration, valid ad-hoc modifications to a trial may be acceptable given those modifications do not compromise the validity, integrity, and feasibility of the trial, and appropriate communication with the regulatory agency to obtain confirmation in advance is required.

In addition, making certain modifications to an ongoing clinical trial based on external data is not considered an adaptive modification, but should be reflected in protocol amendments, which are communicated to the regulatory agency in a timely manner. There are many situations where the protocol is amended based on external data: for example, a trial in which the drug is too toxic for patients who are marker negative or newly completed trial(s) of drug(s) in the same class demonstrated that the effect is only in marker positive patients, thus, the target population needs to be modified to marker positive patients only; newly completed trial(s) of drug(s) in the same class demonstrated that the choice of the primary endpoint is not appropriate, or newly published corresponding guidelines have recommended another primary endpoint definition, which requires modification of the primary endpoint; change in the standard of care treatment requires modification of the control group treatment; or a trial needs to be terminated early because it cannot continue enrolling patients. The sponsor should pay particular attention that these modifications are based on external data only, but not on results from the ongoing trial itself.

## Reference

1. CDE.Guidelines for the Establishment and Implementation of the Clinical Trial Data Monitoring Committee.2020
2. Bauer P, Köhne K. Evaluation of experiments with adaptive interim analyses.*Biometrics* 1994;50: 1029 – 1041, rectionin*Biometrics*1996;52: 380.
3. Bauer P, Kieser M. Combining differentiated phases in the development of medical treatments with in a single trial.*Statistics in Medicine*1999;18: 1833-1848.
4. Bauer P, Posch M. Letter to the Editor.Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections by H. Schaefer and H.- H. Mueller, *Statistics in Medicine* 2001;20: 3741 – 3751. *Statistics in Medicine*2004;23: 1333 – 1335.
5. Chang M. Adaptive design method based on sum of p-values.*Statistics in Medicine* 2007;26: 2772 – 2784.
6. Chen C, Li X, Li W, Beckman RA.Adaptive Expansion of Biomarker Populations in Phase 3 Clinical Trials.*Contemporary Clinical Trials*2018;71: 18-85.
7. ChenJYH, DeMets DL, Lan GKK.Increasing the sample size when the unblinded interim results is proming. *Statistics in Medicine* 2004;23: 1023-1038.
8. ChowSC, Chang M. Adaptive Design Methods in Clinical Trials.*CRC Press*;2nd edition 2011.
9. Chow SC, Shao J, Wang H, Locknygina Y. Sample size calculations in clinical research. *Chapman & Hall/CRC, 3<sup>rd</sup>Edition*2018.
10. Cui L, Hung HMJ, Wang SJ.Modification of sample size in group sequential clinical trials.*Biometrics*1999;55: 853-857.
11. EMA Scientific Guidelines: Reflection paper on methodological issues in conventional clinical trials planned with an adaptive design.<https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines>;2007.
12. FDA Drug Guidance for Industry.Adaptive Designs for clinical trials of drugs and biologics. <https://www.fda.gov/drugs/guidances-drugs/all-guidances-drugs>2018.
13. Friede T, Kieser M. Sample size receiving in internal pilot studies: a review.*Biological Journal*2006;48: 537 – 555.
14. Friede T, Parsons N, Stallard N. A conditional error function approach for subgroup selection in adaptive clinical trials.*Statistics in Medicine*2012;31: 4309 – 4320.
15. Friede T, Stallard Nigel.A component of methods for adaptive treatment selection.*Biological Journal* 2008;50: 767 – 781.
16. Gould AL.Interim analyses for monitoring clinical trials at do not material effect the Type I error rate.*Statistics in Medicine*1992;14: 1039-1051.
17. HochbergY.A sharper Bonferroni procedure for multiple tests of signature.*Biometrika*1988;75: 800 – 802.
18. Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using related survival endpoints.*Pharmaceutical Statistics*2011;10: 347 – 356.
19. Lan KG, DeMetsDL.Discrete Sequential Boundaries for Clinical Trials. *Biometrika*1983;70: 659 – 663.
20. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials.*Biometrics*1999;55: 1286 – 1290.
21. Maca J, Bhattacharya S, Dragalin S.et al.Adaptive Seamless Phase II/III Designs Background, Operational Aspects, and Samples.*Drug Information Journal* 2006;40: 463-474.
22. Marcus R, Peritz E, Gabriel KR.On closed testing procedures with special reference to ordered analysis of variance.*Biometrika*1976;63: 655 – 660.
23. Müller HH, Schäfer H. Adaptive group sequential designs for clinical trials: Combining the advances of adaptive and of classical group sequential procedures.*Biometrics*2001;57: 886 – 891.
24. 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.
25. O'Brien PC, Fleming TR.A Multiple Testing Procedure for Clinical Trials.*Biometrics*1979;549 – 556.
26. Pocock SJ, Group Sequential Methods in the Design and Analysis of Clinical Trials.*Biometrika*1977;64: 191 – 199.
27. Proschan MA, Hunsberger SA.Designed extension of studies based on conditional power.*Biometrics* 1995;51: 1315 – 1324.

28. Rosenblum M, Van Der Laan MJ. Optimizing randomized trial designs to distinguish which subpopulation from treatment. *Biometrika* 2011;98: 845-860.
29. Shih WJ. Sample size re-investigation – a journey for a decade. *Statistics in Medicine* 2001;20: 515-518.
30. Shih WJ, Li G, Wang Y. Methods for flexible sample-size design in clinical trials: Likelihood, weighted, dual test, and promising zone applications. *Contemporary Clinical Trials* 2016;47: 40-48.
31. Simes RJ. An advanced Bonferroni procedure for multiple tests of signature. *Biometrika* 1986;73: 751 – 754.
32. Stallard N, Hamborg T, Parsons N, Friede T. Adaptive designs for congenital clinical trials with subgroup selection. *Journal of Biopharmaceutical Statistics* 2014;24: 168 – 187.
33. Wang SJ, Hung HMJ, O'Neill RT. Adaptive patient design design in therapeutic trials. *Biological Journal* 2009;51: 358 – 374.
34. Wassmer G, Brannath W. Group sequential and continuous adaptive designs in clinical trials. *Springer*; 2016.
35. Wu PS, Lin M, Chow SC. On sample size investigation and re-investigation adjustment for variation in congenital trials. *Journal of Biopharmaceutical Statistics* 2016;26: 44-54.
36. Zhang JJ, Blumenthal G, He K, Tang S, Cortazar P, Sridhara R. Overestimation of the effect size in group sequential trials. *Clinical Cancer Research* 2012;18: 4872-4876.



ARI  
三方协调委员会

## Appendix 1 Glossary

**Interim Analysis:** Any analysis conducted using cumulative data during the trial. One or more interim analyses could be planned within the same trial.

**Adaptive design:** A clinical trial design in which the trial design is to be modified based on the cumulative data during the trial in the interim analysis, according to a prespecified plan. The modification is also referred as adaptive modification. The adaptive modification plan must be prespecified in the trial protocol and statistical analysis plan before the start of clinical trials.

**Validity:** The statistical analysis method used does not bias the estimation of efficacy. The justification for the trial concerns the credibility, interpretability and persuasiveness of the trial results.

**Integrity:** The integrity of an adaptive trial refers to a good control over the bias introduced by the trial operation. Maintaining the integrity of the trial means that modifications should be based on a predefined protocol and keep blinded of the interim analysis results to minimize operation bias.

**Feasibility:** The feasibility of an adaptive trial is about whether the adaptive modification of the trial can be implemented in practice.

**Group sequential design:** A design in which one or more interim analyses are pre-planned to be conducted during the trial, and the decision for the rest of the trial will be made based on the results of each interim analysis.

**Blinded/Non-comparative analysis:** An interim analysis in which the actual treatment group information is not used, or although the actual treatment group information is known, no analysis involving comparisons between groups will be performed, such as the analysis in the interim to pool data from the two treatment groups.

**Non-blinded/comparative analysis:** Refers to the analysis which uses treatment group information (including the actual name of each group or group code) in the interim analysis. The analysis involves comparison between groups.

**Two-stage seamless adaptive design:** A design which divide the trial into two stages. An interim analysis is performed at the end of Stage 1, after which the Stage 2 trial may be adaptively modified based on the prespecified criteria.

**Adaptive enrichment design:** The target population in Stage 2 will be adaptively modified according to the prespecified criteria based on the interim analysis results, following the complete of Stage 1 in the trial.

**Master protocol design:** A design in which a clinical trial has a master protocol combined with multiple sub-protocols. Multiple sub-protocols can test the clinical efficacy of a drug on multiple diseases at the same time, or test the clinical efficacy of multiple drugs on one disease at the same time, or test clinical effect of multiple drugs on multiple diseases at the same time.

**Master protocol with adaptive designs:** A design with one or more adaptive modifications to be included in the master protocol.

**Multiple adaptive design:** A design that uses two or more adaptive modification methods in a trial.

**Bayesian method:** A class of statistical methods that combine information/data from previous trials summarized by a distribution function (prior distribution) with data from current trial, to create a new distribution function (posterior distribution) that summarizes all the information/data according to the Bayesian principle, and draw statistical inferences based on the posterior distribution function.

**Bayesian adaptive design:** A trial design that uses Bayesian approaches and contains adaptive modifications.

**Simulation-based adaptive design:** A design that will assess of the validity of the

statistical inference in adaptive trials, based on simulation methods.





## **Appendix 2 Examples of Adaptive Designs**

The study designs, statistical methodologies, and specific settings of parameters mentioned in the examples are aimed to give a brief description of study design elements, which are not applicable and comprehensive universally.

### **Example 1: Group Sequential Design**

Imagine there is a multi-center, randomized, double-blind, parallel-group, active-controlled, phase 3 superiority clinical trial. The target is to verify the efficacy of an investigational product in improving symptoms for some indication. The primary endpoint is change from baseline after 52 weeks of treatment in a continuous variable with an approximate normal distribution assumption.

The trial uses a group sequential design. The planned interim analyses are: the first interim analysis (safety analysis) is to be performed when  $n_1$  subjects are enrolled. The second interim analysis (futility analysis) is to be performed when  $n_2$  subjects are enrolled. The third interim analysis (efficacy or futility analysis) is to be performed when  $n_3$  subjects are enrolled.

Based on the above design, with the expected efficacy parameters (such as mean and common standard deviation of the change between two groups), test power (such as 90%), the proportion of each group assigned, the overall type I error rate (e.g., 1-sided 0.025) and corresponding control method, the total sample size needed in this trial can be calculated, and the decision strategy for the three interim analyses is made as follows:

- 1) In the first interim analysis, the independent Data Monitoring Committee (DMC) will perform a comprehensive analysis on the safety data, and could make recommendations to terminate the trial if DMC considers there are serious safety

issues for the study drug. This interim analysis is for safety purpose only, therefore, no need to adjust  $\alpha$ .

2) In the second interim analysis, the primary endpoint of change in the control group is evaluated for superiority compared with the test group, and if the control group is superior, the trial will be terminated early due to futility. This interim analysis only examines futility and does not include efficacy assessment, therefore, no need to adjust  $\alpha$ .

3) In the third interim analysis, p-value from the statistical testing of efficacy of the investigational product will be compared with two boundaries, which are decided based on a certain  $\alpha$  and  $\beta$  spending function. If the p-value is less than the lower bound, the trial will be terminated early with efficacy; if the p-value is greater than the upper bound, DMC could make proposals either to early terminate or to continue the study after a comprehensive consideration; the study will be continued if the p-value is between the upper and lower bound.

### **Example 2: Blinded Sample Size Re-estimation**

Imagine a randomized, double-blind, placebo-controlled, parallel-group, phase 3 superiority clinical trial. The primary endpoint is change from baseline at Visit 4 in a scale score, following an approximate normal distribution assumption. The overall mean of difference in the primary endpoint between test group and control group is 6.0, the standard deviation 10.0, the nominal test level is set to one-sided 0.025, and the test power is 90%. The two groups are designed to be balanced, and the total sample size is 120 (60 subjects in each group). If the dropout rate is 20%, the total sample size for the original design will be 150. A sample size re-estimation is planned in the interim analysis, given that the original setting of standard deviation 10.0 may not be correct.

The interim analysis plan is to perform the interim analysis in a blinded manner, the pooled standard deviation of the cumulative data will be calculated when approximately 50% of the original sample size complete or discontinue the study. If the pooled standard deviation is greater than 10.0 in the original assumption, the final total sample size will be re-estimated based on this value, and other parameters will keep consistent with the original ones; if the pooled standard deviation is less than 10.0, the sample size remains the same as 150. Continue to enroll subjects till the study end according to the re-estimated sample size in the interim analysis or the original one. No adjustment for type I error rate. Consider the case when pooled standard deviation is 13.66. If standard deviation is 13.66, mean is 6.00 same as original design, the sample size recalculated will be 220. The total sample size will be 275 with regards to a 20% dropout rate.

### **Example 3: Unblinded Sample Size Re-estimation**

Imagine a multi-center, randomized, double-blind, active-controlled, parallel-group, phase 3 superiority clinical trial. The primary endpoint is change from baseline at Week 24 in a standard score with an approximate normal distribution assumption. Z-Test is used in the efficacy analysis. According to the parameters expected in the trial (e.g., the difference between sample means of the two groups is  $\delta_0$ , standard deviation  $\sigma=1$ ) and other required elements (e.g.,  $\alpha=0.025$ , power  $1-\beta=90\%$ ), the initial total sample size is determined to be  $N$ .

Assume an interim analysis and sample size re-estimation is to be conducted when  $n_1$  subjects complete. Let  $n_2=N-n_1$  be the sample size in Stage 2 under the initial design and  $n_2^*$  be the sample size in Stage 2 based on the results of between-group comparison in the interim analysis. Consequently, the sample size to be added in Stage 2 is  $n_2^*-n_2$ , and  $N^*$  is the total sample size after sample size increase.

The determination of  $n_2^*$  requires a reasonable adaptive modification method. Since different adaptive modification methods have their own advantages and disadvantages, and some methods may be the special cases of other methods given some conditions, it is difficult to make a clear statement about how to choose these methods. The choice of methods should consider the trial objectives, assumptions, and analytical methods. Simulation methods may also be a choice. The key design elements of the methods commonly used are briefly described as below. It should be noted that due to the limitation in research & development cost, plenty of methods will set an upper limit to the sample size in calculation, while many methods will take minimal value of efficacy to be tested with clinical meaningfulness into account during calculation. The following introduction disregards these two factors.

1) Promising zone method: conditional power  $CP(N, z_1)$  is calculated based on the interim analysis results, and is divided into three zones, i.e.  $(0, c_1]$ ,  $(c_1, c_u)$  and  $[c_u, 1)$ .  $z_1$  is the Z-statistics obtained from the data of  $n_1$  subjects in Stage 1. If  $CP(N, z_1) \leq c_1$ , the study would be terminated; if  $CP(N, z_1) \geq c_u$ , it would be continued; if  $c_1 < CP(N, z_1) < c_u$ , the sample size will be re-estimated: when  $CP(N, z_1) > 50\%$ , the re-estimation of sample size will not increase type I error rate. As to the sample size re-estimation, the efficacy difference between two groups obtained from the interim analysis can be brought into the sample size calculation formula to calculate the new sample size  $N^*$  which could meet the criteria of  $CP(N^*, z_1) = 1 - \beta$ . The Z test statistics from the adjusted sample size  $N^*$  will be compared with  $z_{1-\alpha}$  (no adjustment required) to evaluate efficacy in the final analysis.

2) Weighted statistics method: it can also be regarded as an inverse normal distribution combination function method, but simpler. Based on the interim analysis results, the adjusted sample size  $N^*$  is calculated based on certain criteria (e.g., conditional test power). Let  $w_1 = \sqrt{n_1/(n_1 + n_2)}$ ,  $w_2 = \sqrt{n_2/(n_1 + n_2)}$ , and the

final test statistic  $Z^* = w_1 Z_1 + w_2 Z_2^*$  will be compared with  $z_{1-\alpha}$  (no adjustment required). Here,  $Z^*$  follows a standard normal distribution,  $Z_1$  is the Z-statistics based on data from  $n_1$  subjects in Stage 1 only, and  $Z_2^*$  is the Z-statistics from  $n_2^*$  subjects in Stage 2 after adjustment. It should be noted that the above weight calculation is based on  $n_1$  and  $n_2$  in the original trial design. Yet with only  $n_1$  specified could we also apply this kind of weighted statistic methods, without the initial total sample size  $N$ .

3) Maximum Likelihood Ratio Method: set the target value to be achieved  $CP(N^*, z_1)$ . Accordingly adjust the critical point of the rejection region  $c$ , which is given by the formula  $1 - \alpha = \int_{-\infty}^k \Phi \left[ \frac{c\sqrt{n_1+n_2^*} - z_1\sqrt{n_1}}{\sqrt{n_2^*}} \right] \phi(z_1) dz_1$  (here  $\Phi(\cdot)$  and  $\phi(\cdot)$  represent the cumulative distribution function and probability density function of the standard normal distribution respectively). Here  $z_1 < k$ , and  $k$  is decided by a selected  $\alpha$  spending function (in case of an efficacy testing). The Z-statistic will be calculated based on the adjusted sample size in the two stages, and compared with the adjusted cutoff  $c$  in the final analysis.

4) Conditional error function method: select an incremental function  $A(z)$  meeting certain criteria. Let  $z_A = \Phi^{-1}(1 - A(z_1))$ , and  $n_2^* = 2(z_A + z_\beta)^2 / \delta_1^2$ , where  $\delta_1$  is the estimated value of efficacy in the interim analysis, while  $z_\beta$  is the  $\beta$  quantile of the standard normal distribution, and  $n_2^* - n_2$  is the sample size that needs to be added. In the final analysis,  $Z_2^*$  which is calculated based on the adjusted sample size in Stage 2, will be compared with  $(z_1^2 + (z_A + z_\beta)z_A) / \sqrt{z_1^2 + (z_A + z_\beta)^2}$  to evaluate efficacy.

5) Method based on sum of p-values: Let  $N^* = |\delta_0 / \delta_1| 2N$ . Here  $\delta_1$  is the estimated value of efficacy based on the data from interim analysis. Calculate p-

values in the two stages as  $p_1$  and  $p_2$  respectively. Compare  $p_1 + p_2$  with 0.2236 for efficacy evaluation. This method has a restriction that the efficacy estimates  $\delta_1$  obtained from interim analysis should not be in the opposite direction with the original one  $\delta_0$ , and this method can't reduce sample size.

#### **Example 4: Two-stage Seamless Phase II/III Adaptive Design**

Imagine there is a multi-center, randomized, double-blind, active-controlled, parallel-group, superiority study to assess the improvement of a new drug in some symptom. The primary endpoint is change from baseline at Week 8 in a score, which is assumed to follow a normal distribution approximately. It is planned to use a two-stage seamless phase II/III adaptive design. The phase II trial has two drug groups (high dose and low dose) and one control group with a randomization ratio of 1:1:1. Efficacy, defined as the difference of mean improvement in the score between subjects in two groups, is expected to be  $\delta$ . Let one-sided  $\alpha=0.025$  and power  $1-\beta$  (e.g., 90%). Sample size  $N$  is calculated for the comparison between one dose group and the control group. When  $3n_1$  subjects finish their 8-week follow up, Stage 1 ends. Total sample Size is  $N+n_1$  (number of subjects in the two comparison groups  $N$  + number of subjects in the de-selected dose group in Stage 1  $n_1$ ).

For Stage 1, let  $p_{11}$  and  $p_{12}$  be the p-value of Z-test between low dose group vs. control group (null hypothesis  $H_{011}$ : there is no difference between the low dose group and the control group) and high dose group vs. control group (null hypothesis  $H_{012}$ : there is no difference between the high dose group and the control group), respectively. Closed testing procedure and Hochberg procedure is used to adjust multiplicity. The p-value of the test of no difference between either the low or high dose group and the control group ( $H_{011} \cap H_{012}$ ) is  $p_{int1} = \min[2 * \min(p_{11}, p_{12}), \max(p_{11}, p_{12})]$ . The p-value from efficacy comparisons between the selected dose

group in Stage 1 and the control group will be  $p_1 = \max(p_{int1}, \min(p_{11}, p_{12}))$  after multiplicity adjustment.

The number of subjects to be enrolled in Stage 2 should be the original sample size of the two groups  $N$  minus the number of subjects enrolled in the two groups in Stage 1  $2n_1$ , that is  $N - 2n_1$ . Enrolled subjects are randomly assigned to the selected dose group and the control group. When the study ends,  $p$ -value in each stage will be calculated based on the data from Stage 1 and Stage 2 respectively, denoted as  $p_1$  and  $q$ . Inverse normal combination test is used, and  $C(p_1, q) = 1 - \Phi[w_1 \Phi^{-1}(1 - p_1) + w_2 \Phi^{-1}(1 - q)]$  is calculated to be compared with  $\alpha = 0.025$  for efficacy evaluation. Weight  $w_1$  is the square root of ratio between number of subjects in three groups in Stage 1 and total number of subjects enrolled, i.e.,  $w_1 = \sqrt{3n_1 / (N + n_1)}$ . Weight  $w_2$  is the square root of ratio between number of subjects in two groups in Stage 2 and total number of subjects enrolled, i.e.,  $w_2 = \sqrt{(N - 2n_1) / (N + n_1)}$ .

### **Example 5: Adaptive Enrichment Design**

Imagine there is a multi-center, randomized, double-blind, active-controlled, parallel-group, two-stage superiority clinical trial. The primary endpoint is overall survival (OS), with the secondary endpoint progression-free survival (PFS). Assume the hazard ratio is 0.75 in the overall population HR (F), and 0.55 in the positive subgroup HR (S). Use 1-sided test level 0.025, test power 90%, and other necessary parameters to calculate the total number of deaths required in the overall population, denoted as  $N_0$ .

The trial is planned to select the target population in the interim analysis when 40% subjects are enrolled. The duration of overall survival is quite long in this study, hence the selection of target population in the interim analysis is based on PFS. The decision strategy is: ① If estimated HR (F) < 0.85 and HR (S) < 0.65, the trial will be

continued in both the positive subgroup and the overall population in Stage 2; ② If  $HR(F) \geq 0.85$  and  $HR(S) < 0.65$ , only subjects in the positive subgroup will be enrolled in Stage 2; ③ If  $HR(F) < 0.85$  and  $HR(S) \geq 0.65$ , the trial will be continued in the overall population in Stage 2, and the positive subgroup will not be analyzed; ④ The trial will be terminated for futility if  $HR(F) \geq 0.85$  and  $HR(S) \geq 0.65$ .

Let  $p_{1F}$ ,  $p_{1S}$  and  $p_{1FS}$ ,  $p_{2F}$ ,  $p_{2S}$  and  $p_{2FS}$  be the p-value of null hypothesis  $H_{0F}$ ,  $H_{0S}$  and  $H_{0F} \cap H_{0S}$  in two stages, respectively. The overall type I error rate of combined test  $H_{0F} \cap H_{0S}$  will be controlled at a 1-sided level of 0.025 by applying the inverse normal combination function and closed method. If Simes method is adopted, the p-value of test ( $H_{0F} \cap H_{0S}$ ) will be  $p_{FS} = \min(2\min(p_F, p_S), \max(p_F, p_S))$ . Let  $C(p, q) = 1 - \Phi[w_1 \Phi^{-1}(1 - p) + w_2 \Phi^{-1}(1 - q)]$ , in which,  $w_1 = \sqrt{\frac{n_1}{n_1 + n_2}}$ ,  $w_2 = \sqrt{\frac{n_2}{n_1 + n_2}}$ ,  $n_1$  and  $n_2$  represents the number of deaths in two stages respectively. The p-value of test  $H_{0F} \cap H_{0S}$  in the final analysis is  $C(p_{1FS}, p_{2FS})$ .  $H_{0F} \cap H_{0S}$  would be rejected when  $C(p_{1FS}, p_{2FS}) \leq 0.025$ . In case when only subjects in the positive subgroup are enrolled in Stage 2, and  $C(p_{1FS}, p_{2S}) \leq 0.025$ ,  $H_{0F} \cap H_{0S}$  would be rejected. Similarly,  $C(p_{1F}, p_{2F})$  and  $C(p_{1S}, p_{2S})$  could be used to test  $H_{0F}$  and  $H_{0S}$ . Based on the closed method,  $H_{0F}$  will be rejected eventually if  $H_{0F} \cap H_{0S}$  and  $H_{0F}$  could be rejected at the same time, and  $H_{0S}$  will be rejected eventually if  $H_{0F} \cap H_{0S}$  and  $H_{0S}$  could be rejected at the same time.

As to this trial, if the test in Stage 1 is based on PFS and the test in Stage 2 is based on OS, it will be difficult to interpret the meaning of p-values of two stages in the final analysis. Therefore, in this trial design, decision making of Stage 1 is based on descriptive statistics of PFS, while in final analysis, the p-value of two stages is based on OS. For trials using survival as an endpoint, no matter in which stage the endpoint event occurs for subjects enrolled in Stage 1, it should be included in the



results of Stage 1 in the calculation. Otherwise, the assumption of independence between two stages will no longer be held, and type I error rate will be inflated.

### **Example 6: Adaptive Master Protocol Study**

Imagine there is a superiority clinical trial which is to test a new drug in treating rare cancer patients with BRAF V600E mutation positive. The primary endpoint is the objective response rate confirmed by an independent endpoint committee, and the time to sustainable response is documented. A multi-center, single-arm basket design is adopted. Subjects enrolled in the trial must be in the late stage of the disease and have a central laboratory confirmed BRAF V600E mutations. There are five cohorts, including anaplastic thyroid cancer, biliary tract cancer, gastrointestinal stromal tumor, hairy cell leukemia, and small intestine adenocarcinoma.

Although all subjects are enrolled under the same protocol in this trial, each of the 5 cohorts will be considered as an independent trial with separate results to support the submission of corresponding cohorts. Because the purpose of the trial is to support the application of a new drug, the sample size must be determined in advance, and the sample size required for each cohort should be calculated separately according to the decision rule of superiority. As to the consideration to combine data from two or more cohorts, due to the lack of sufficient data in this trial to support the investigational product have the same mechanism of action and similar efficacy in patients with BRAF V600E mutation positive, therefore, it is not acceptable to combine data from any two or more cohorts to support the application of a new drug in the corresponding pooled cohorts.