

# **Guideline on Statistical Design of Cancer Clinical Trials (Trial Version)**

*English Translation by: Cong Chen, Kun He*

*Disclaimer: The English is for information only and not an official translation and under any dispute the Chinese will prevail*



**Center for Drug Evaluation, NMPA**

**November 2021**

# Contents

<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>2. EFFICACY ENDPOINTS .....</b>	<b>2</b>
2.1 Overall Survival (OS) .....	2
2.2 Objective Response Rate (ORR) .....	4
2.3 Progression-Free Survival (PFS) .....	5
2.4 Patient-reported Outcome (PRO) .....	9
<b>3. EXPLORATORY TRIAL.....</b>	<b>10</b>
3.1 Dose-finding Design .....	10
3.2 Single Arm Trial and FIH Cohort Expansion.....	11
<b>4. CONFIRMATORY TRIAL.....</b>	<b>13</b>
4.1 General Considerations.....	13
4.2 Trial Design .....	14
4.2.1 Group Sequential Design.....	15
4.2.2 Two-stage Adaptive Design .....	15
4.2.3 Design with Biomarkers .....	17
4.2.4 Master Protocols Design.....	18
<b>REFERENCES.....</b>	<b>21</b>
<b>Appendix 1: Abbreviations .....</b>	<b>错误!未定义书签。</b>

# **Guideline on Statistical Design of Cancer Clinical Trials**

## **1. INTRODUCTION**

Oncology drug development is similar to other disease areas, before any clinical trial is carried out, there should be sufficient scientific evidence based on preclinical experiments and previous human trials to indicate that the experimental drug is acceptably safe at the dose(s) in the target patient population. The primary objectives of a clinical trial are to ask important questions pertinent to the drug development and answer them with appropriate design and analysis of the study. Randomized controlled trial (RCT) design is the gold standard in drug development. The reliability of a drug's effectiveness and safety based on a non-randomized trial will not be as strong as those based on a randomized trial.

Since cancer is life-threatening generally with an unmet medical need, clinical research and development of anti-cancer drugs have their uniqueness. For example, early clinical trials use patients as research subjects, rather than healthy subjects; in some cases, the results of single-arm trials are used for regulatory registration. For different cancer indications, sponsors should have different clinical research and development strategies, and have different goals and roles for exploratory trials and confirmatory trials in research and development plans. Clinical

trial design is one of the most important factors that determine the success of drug development. Good experimental designs will help to achieve the objective of the trials and improve the efficiency of research and development.

Innovative clinical trial designs and methods are constantly emerging. Through continuous practice, the experience of anti-cancer drug development and regulatory review is gradually growing too. The purpose of this guideline is to provide scientific advice on key statistical issues in the design of anti-cancer drug trials, and provide points of considerations for sponsors during clinical research and development of anti-cancer drugs. This guiding principle only represents current views and knowledge. It will continue to be revised and improved after more experience accumulated.

## **2. EFFICACY ENDPOINTS**

The most used efficacy endpoints in oncology clinical trials include overall survival (OS), objective response rate (ORR), and progression-free survival (PFS), etc.

### **2.1 Overall Survival (OS)**

Defined as the time from randomization (or treatment start in single-arm trials) until death from any cause. OS is objective, precise, and easy to measure, and is generally regarded as the most reliable endpoint for measuring clinical benefit in randomized clinical trials in oncology.

Generally, OS should be analyzed in the intention-to-treat (ITT) population. The ITT analysis includes all study participants who are randomized

according to a pre-specified study protocol or all study participants who received any amount of drugs in single-arm trials regardless of noncompliance, protocol deviations, withdrawal, and any events after randomization or treatment started in single-arm trials. Because participants who are lost to follow-up tend to be at higher risk of death and bias arises when censoring time or proportion is unbalanced between two treatment arms, assessment of imbalance on censoring pattern should be evaluated. Every effort must be made to ensure that all study participants, regardless of which treatment arms they are assigned to, have the most up-to-date survival information at the time of analysis. Demonstration of a statistically significant improvement in OS can be considered clinically meaningful if the safety profile is acceptable, and can be used to support a regulatory approval of the experimental drug.

The log-rank test is normally used for hypothesis testing of OS, while the Cox-regression model is often used for estimation of the relative treatment effect (hazard ratio). Survival probabilities are usually calculated by the Kaplan-Meier method and presented in survival curves. The log-rank statistic gives all events the same weight, regardless of the time at which an event occurs. If a stratified log-rank test is used, the stratification factors used in the log-rank test must be pre-specified from the stratification factors during the randomization. Alternative weighting methods may be also considered if the proportional hazards assumptions do not hold. However, since mild deviation from proportional hazards is commonly seen and it is extremely hard to predict the pattern of hazard ratio over time

based on past clinical experience, before proposing a different weighting method, the pros and cons of the proposed method should be carefully evaluated and also be discussed with the Regulatory Agency.

Cross-trial comparisons of OS are unreliable, as different trials may differ in terms of patient selection, choice of standard of care (SOC) or best supportive care (BSC), etc. Therefore, it should be cautious to use and interpret OS results in single-arm trials.

## **2.2 Objective Response Rate (ORR)**

For many cancer types, radiographic tumor assessments can be directly used to evaluate the disease, and the treatment strategies are also based on the tumor assessment result and clinical symptoms. Defined according to the commonly accepted response evaluation criteria (e.g., RECIST version 1.1 for solid tumors), objective response rate (ORR) refers to the proportion of study participants with tumor size reduction of a pre-defined amount and for certain minimum time duration. ORR is the most popular endpoint based on tumor measures. A response for solid tumors can be a complete response (CR) or a partial response (PR), and there are other evaluation criteria for non-solid tumors. ORR alone may not adequately describe the anti-tumor activities of an experimental drug, and therefore endpoints like time to response and duration of response (DOR) (i.e., the time from documentation of tumor response to disease progression or death) is also routinely analyzed simultaneously with ORR. For drugs that provide clinical benefits via disease stabilization, disease control rate (DCR) may also be analyzed. DCR captures not only subjects with responses, but also

cases where disease remains at a stable status for a certain duration. Tumor size change over time relative to the baseline is routinely treated as a continuous variable and displayed in a water-fall plot to assess anti-tumor activity.

For trials (single-arm or randomized) intended for registration, tumor measurement and response assessment are usually done by a Blinded Independent Central Review (BICR). If ORR is the primary endpoint, initial response generally requires confirmation in subsequent assessments. In clinical practice, the decision on whether to continue treating a study participant is made by the investigator, and hence discordance in response assessment between individual investigators and BICR may introduce bias in ORR analysis.

Same as for OS, ORR should generally be analyzed in the ITT population. In the ITT analysis, study participants who discontinued before the first tumor assessment should be considered as non-responders regardless of the reason for discontinuation. This way, the comparison of ORR with historical control is more reliable as the latter is often based on the ITT population in a confirmatory trial. Similarly, response assessment should be based on the same response criteria as historical control in order to make a fair comparison. If different response evaluation criteria are used, any impacts due to using different response evaluation criteria should be evaluated.

### **2.3 Progression-Free Survival (PFS)**

PFS is defined as the time from randomization (or treatment start in single-arm trials) until tumor progression or death, whichever occurs first. Similar to PFS, disease-free survival (DFS) is defined as the time from randomization (or treatment start in single-arm trials) until disease recurrence or death due to any cause, which is often used in the adjuvant setting after definitive surgery or radiotherapy, and event-free survival (EFS), defined as the time from randomization (or treatment start in single-arm trials) to the first occurrence of any of the following events: progression of disease that precludes definitive surgery, local or distant recurrence, death due to any cause and so on, which is often used in the neoadjuvant setting prior to definitive surgery or radiotherapy. Other similar endpoints include time to progression (TTP) and time to treatment failure (TTF). Both TTP and TTF are often not accepted as the primary evidence to support the study conclusion and are often used as supportive evidence for PFS.

The precise definition of tumor progression is important for efficacy endpoints based on tumor measurement and should be carefully and prospectively specified in the protocol. Similar to ORR, the definition of progression should follow established response evaluation criteria. PFS is hardly interpretable in single-arm trials because some trial participants can have long stable disease even without active treatment. Therefore, registration trials with PFS as the primary endpoint must have a control arm. In a randomized, double-blinded, active-controlled trial, whether PFS should be based on a BICR assessment depends on safety profile of the



study drug and practice of tumor assessment for the particular disease. When BICR is deemed optional, the tumor images should be archived for future auditing and inspection purpose.

Interval censoring, which refers to the situation that disease progression occurs in between of two tumor assessments, is a challenge in the PFS analysis. At the time of tumor assessment, a decision that disease has progressed according to established criteria only means that progression has occurred sometime between the last assessment and the present one. The consequence is that estimation of PFS depends on the assessment schedule, so that comparison of median PFS between treatment arms will be biased if the tumor assessment schedules are different. While interval-censored data analysis methods can account for difference in individual assessment schedules to some extent, it is strongly recommended that the tumor assessment schedules are kept identical between treatment groups to improve accuracy of estimation and avoid complexities in analysis and interpretation of the data.

Another more challenging issue with PFS analysis is informative censoring. Suspicion of informative censoring arises from various scenarios, with those commonly seen include: 1. despite of no evidence of disease progression, a trial participant may have violated the protocol by taking other anti-cancer treatment during the trial; 2. a trial participant may discontinue treatment based on a progression concluded by a local investigator but is then overwritten by the BICR; 3. a trial participant may discontinue treatment due to toxicity without any evidence of disease

progression. The true outcome would be unknown if tumor assessments have been terminated; 4. a trial participant's actual assessment time may deviate from the scheduled time due to the worsening of the underlying disease.

PFS should be analyzed following the ITT principle. If progression is detected during an unscheduled assessment, the date of progression should rather be recorded based on the documented time of progression instead of the scheduled time of assessment. At the time of the analysis, every effort must be made to ensure all study participants, including those discontinued treatment without documented progression, will have the most up-to-date tumor assessment information. Time-to-censoring analysis may help reveal follow-up imbalance between two treatment arms. Discordance in disease progression assessment between investigators and BICR is an important topic in PFS analysis. An analysis of discordance should be routinely conducted at the time of analysis to investigate if there is any imbalance between treatment arms. Standard statistical methods for survival data all heavily rely on validity of the non-informative censoring assumption. When informative censoring is suspected, the most practical suggestion is to perform sensitivity analyses under assumptions that are likely consistent with the truth. For example, in the first two scenarios mentioned above where informative censoring could occur, a sensitivity analysis can be considered that modifies the definition of progression to more closely reflect clinical judgment of treatment failure.

PFS is usually treated as a right-censored time-to-event variable and is analyzed by the same methods as for OS. However, the estimates of median PFS may not reflect the real treatment effect in some trials. For example, two treatment arms may have similar median PFS despite of large treatment effect as reflected by hazard ratio. The same assessment schedules that study participants follow also give rise to tied event times. Exact (or approximately exact) method for tie-handling is recommended in estimation of treatment effect under the Cox-regression model. When calculating the sample size, attention should also be paid to the loss of information due to interval-censoring, because the conventional practice that treats PFS as a right-censored time-to-event variable in such calculation may overestimate the study power. The issue in PFS is more pronounced when assessment is less frequent relative to time to disease progression.

#### **2.4 Patient-reported Outcome (PRO)**

Patient-reported outcome measures any outcome evaluated directly by the study participant in terms of symptoms, health-related quality of life, health status, adherence to treatment, and satisfaction with treatment. While PRO data have been more often collected in cancer trials, several issues exist in the evaluation of such measurements, such as validity, reliability, and reactivity, etc. Besides, the measurements are easily affected by missing data, which needs to be addressed using appropriate methods. Therefore, rarely, PRO results are used as primary evidence for supporting a regulatory application. To better understand the relevance of the trial

outcomes, the relationship between PRO and other efficacy endpoints should be explored.

### **3. EXPLORATORY TRIAL**

#### **3.1 Dose-finding Design**

Phase 1 cancer clinical trials are often the studies in which the experimental drug is tested first time in human (first-in-human, FIH). The guiding principle for dose escalation in Phase 1 cancer clinical trials is to avoid unnecessary exposure of trial participants to subtherapeutic or over therapeutic doses of the experimental drug (i.e., to treat as many participants as possible within the therapeutic dose range) while preserving safety and maintaining rapid accrual. Formal dose escalation methods for Phase 1 cancer clinical trials fall into two broad classes: the rule-based designs including the traditional 3+3 design and its variations that are not supported by any statistical modeling, or the model-based designs such as the continuous reassessment method (CRM). Emerging hybrid methods such as the modified toxicity probability interval (mTPI) design and the Bayesian optimal interval (BOIN) design are model-based but allow pre-specification of dose escalation rules. These methods are easy to implement and have the flexibility to choose the target toxicity rate and cohort size, and also have comparable performance with the model-based designs.

In order to minimize the number of participants treated at potentially subtherapeutic doses, Phase 1 dose-finding may start with an accelerated titration, which usually enrolls 1-3 participants at each dose level and ends

with the occurrence of a Grade 2 or higher non-disease related toxicity. After the accelerated titration part ends, dose-finding will proceed with a formal dose escalation method. In certain circumstances, intra-patient dose escalation (i.e., a participant is treated at a higher dose level in subsequent cycles than in the first cycle) may also be considered, but the safety and tolerability data beyond the first cycle are often hard to interpret. For the determination of a dose as a candidate of recommended phase 2 dose (RP2D), adequate number of participants should be treated at this dose.

### **3.2 Single Arm Trial and FIH Cohort Expansion**

In contemporary oncology drug development, a single-arm trial is often conducted in one or more tumor indications after the dose-finding ends to further explore the safety of drugs and preliminarily investigate the efficacy. The tumor indication cohorts may be formed by different tumor types in same line of therapy, different lines of therapies of same tumor type, or a combination of both. Study participants in a cohort may be treated with the experimental drug as a monotherapy or as a combination therapy (e.g., with a SOC or another experimental drug).

The study protocol for such a single-arm trial should contain adequate information justifying the planned sample size based on the cohort objectives and specify the magnitude of anti-tumor activity that would warrant further evaluation of the drug. In a nonrandomized cohort, assessment of anti-tumor activity is generally determined by a multi-stage design to limit exposure of additional patients to an ineffective drug. Details on whether accrual will be paused and the minimum duration of

follow-up time for participants for an interim analysis should be specified in the protocol. The need for comparison of safety and anti-tumor activity among different dosing regimens (e.g., two RP2D candidates or monotherapy vs combination therapy) may be addressed in a randomized cohort with more statistical rigor.

By the time a single-arm trial design is initiated, there may not be adequate data on the metabolism and pharmacokinetics of the experimental drug or enough safety assessment. Rapid enrollment, especially in case of exciting preliminary signal, potentially exposes large number of study participants to a drug with unknown efficacy and unclear toxicity characteristics. To mitigate such a risk and to protect study participants, it is imperative that sponsors establish an infrastructure to streamline trial logistics, facilitate data collection, and incorporate plans to rapidly assess emerging data in real-time and to disseminate interim results to investigators and institutional review boards (IRBs). According to the results of interim analysis and the pre-specified decision rule, sponsors should pause or terminate the enrollment of cohorts with insufficient antitumor activity or unacceptable safety risk as soon as possible or even terminate the failed trial early.

For expansion cohorts intended for registration, there should be clear distinction between the patient population used for generating the hypothesis on drug activity and the patient population used for confirming this hypothesis. A separated clinical trial is recommended for the purpose of hypothesis testing, especially when the FIH study has undergone

multiple changes on the study population and sample size. Without an active control arm, the trial data need to be very convincing to demonstrate efficacy. When designing a single arm for registration, one should carefully evaluate the prior data and consider an appropriate sample size.

A single-arm trial design is not appropriate for the study of the combination of two novel experimental drugs, unless the contribution of each is well understood and can be separated out.

## **4. CONFIRMATORY TRIAL**

### **4.1 General Considerations**

In designing a confirmatory trial, sponsors should clearly state the targeted treatment effect based on the objective of the trial. The sponsors should also state the patient population, endpoints, treatment plan and possible concomitant events that may affect the estimation of the treatment effect during the trial, such as death, between-group cross-over, etc. The group-level summary statistics, statistical models, and corresponding sensitivity analyses should all be pre-specified.

Although there is a general wish to reduce heterogeneity of study populations in order to increase the study power, restriction of the patient population makes it hard to assess the relevance of the new drug in real world. The choice of control arm should be justified, and in general, it should be selected from best supportive care (BSC) or standard of care (SOC) or investigator's choice as the control.

Double-blinded design is one of the most important methods of controlling bias. When an open-label design has to be used (e.g., due to obvious differences in toxicity profile between study regimens), all conceivable measures must be undertaken to limit the potential bias. Regardless of open-label or double-blinded design, one should consider important prognostic factors and also prognostic covariates potentially impacting the treatment for stratified randomization. Adjusted analyses for covariates should be pre-specified in the protocol or the SAP. When using a predictive biomarker as a stratification factor, the biomarker as well as its cut-point for determining biomarker status (positive or negative) must be pre-specified and the assessment must be based on a validated assay.

The Type I error rate for a confirmatory trial must be strictly controlled under an appropriate level. If the trial objectives involve multiple populations (e.g., a biomarker positive population and an all-comer population), or multiple endpoints (e.g., OS, PFS, and ORR), or an interim analysis plan to terminate the trial early for efficacy, an appropriate method of controlling such multiple comparisons should be pre-specified with details in the protocol or SAP. A trial which may be terminated early for efficacy should consider the adequacy of safety data.

The statistical methods involved in confirmatory trials usually are more complex and difficult, sponsors should communicate with the regulatory agency regarding the statistical methods with their technical details.

## **4.2 Trial Design**



The traditional clinical trial designs used in oncology trials may refer to ICH E9 and other related guidance. Due to better understanding of the disease and faster advance in the field, many innovative designs are used in confirmatory oncology trials, including group sequential design, two-stage adaptive design, design with biomarkers, and master protocols design. These new designs improved the efficiency of the trials tremendously.

#### **4.2.1 Group Sequential Design**

Group sequential designs are routinely used for data monitoring in chronological order or statistical inference of cumulative data. When designing a group sequential trial, the sponsor should carefully consider how many interim analyses are to be planned, when they should be conducted, and which alpha-spending function is appropriate. For trials stopped early for efficacy, sponsors are encouraged to continue follow-up the trial until data maturity to better understand the long-term clinical benefit of the experimental drug.

When the timing of an interim analysis or final analysis is event-driven, the primary dataset should be based on the data cut-off at the time when the target number of events is reached. Every effort should be made to ensure that the data collection and cleaning are completed in a blinded manner before unblinding for the analysis. Data collected afterward will be heavily scrutinized or even dismissed from the analysis due to collection bias.

#### **4.2.2 Two-stage Adaptive Design**

Traditional drug development follows a sequence in which a Phase 2 trial is followed by a Phase 3 trial. Phase 2 trials are used for clinical proof of

concept, dose selection, population selection, or even endpoint selection. The decision to commence Phase 3 is made after Phase 2 data become available. A Phase 3 trial takes time to plan, initiate and implement. Adaptive seamless Phase 2/3 designs are special cases of general two-stage adaptive designs which attempt to eliminate the space between Phase 2 and Phase 3. It can be operationally seamless that excludes Phase 2 participants from the primary analysis, or inferentially seamless that includes Phase 2 participants in the primary analysis. Multiplicity adjustment for Type I error control is not required for the former but may be required for the latter depending on the nature of the adaptation and the hypothesis testing strategy.

Two important factors should be considered before deciding to take a seamless approach as opposed to a sequential approach. First, there should be enough information at the time of seamless transition from Phase 2 to Phase 3 to support a reasonable decision. This often depends on number of participants in the analysis and the usefulness of Phase 2 endpoints for decision making. Second, the operation should be logistically feasible. A seamless design requires expedited data cleaning and analysis as well as quick Phase 3 enrollment. It also requires the timely availability of drug formulation for commercialization. A critical consideration on deciding between an operationally seamless design and an inferentially seamless design is the complexity of adaptive decisions in Phase 2. Unlike for an operationally seamless design, the consistency in trial outcome between Phase 2 and Phase 3 is critical for an inferentially seamless design.

While two-stage adaptive designs have promising potential in accelerating drug development, the advantages and disadvantages of different approaches need to be thoroughly weighed before taking this route. Trial design, operational and statistical issues need to be resolved and discussed with the regulatory agency before starting the trial.

#### **4.2.3 Design with Biomarkers**

To optimize the benefit-risk profile of an experimental drug, it is critical to identify its proper target population. A suitable biomarker may be identified and measured by a variety of different diagnostic approaches (e.g. expression profiling of transcripts, differential antigen expression, genetic diagnostics, including next generation sequencing, etc.). With a multitude of possibilities, it is challenging to determine which biomarkers may be predictive of drug activity and how to set the cut-off value of the biomarker during early development. To minimize selection bias, a training set for biomarker finding and a validation set for biomarker confirmation should be pre-specified to separate study participants into two groups. This hypothesis generation and testing process need to be repeated each time when a new biomarker is investigated. Despite of the investigational rigor, a confirmed predictive biomarker based on a single-arm trial may still be just a prognostic biomarker, in which case a prospective epidemiological study may be conducted to assess the prognostic effect, or a biomarker only predictive of short-term tumor response, in which case a longer follow-up is necessary.

The uncertainty must be accounted for in the design of subsequent confirmatory trials. For example, when it comes down to alpha-allocation between two subpopulations, a step-down approach requires high certainty of the hierarchy, which may not be adequately supported by previous data. In this case, proper alpha-splitting may be preferred. Furthermore, the statistical designs with population selection and expansion can be complicated with Type I error control. The advantages and disadvantages of various design options should be compared and regulatory concerns should be addressed before implementation.

#### **4.2.4 Master Protocols Design**

A trial that tests multiple experimental drugs and/or multiple tumor indications in parallel under a single protocol, without a need to develop new ones for every trial, is called a master protocol. It includes a basket design, an umbrella design, and a platform design.

A trial of an experimental drug that simultaneously investigates multiple tumor indications, in patients with or without biomarker enrichment, is called a basket trial. The primary population of a confirmatory basket trial often includes patients with a unique molecular signature.

The initial selection of tumor indications must be based on significant scientific and clinical evidence so that justification for pooling is on a solid footing, so the risk of trial failure can be reduced. The risk may be further minimized by removal of tumor indications that are less effective from the final pooled analysis based on interim data. Removing the tumor indications with poor treatment benefit based on interim results may

potentially trigger issues with regards to Type I error control, and therefore proper multiplicity adjustment will be required. After removal of ineffective indications, sample size for the remaining tumor cohorts is subject to adjustment to maintain the power of the final pooled analysis. In such a case, the sample size adjustment strategy must be pre-specified and aligned in priori with the regulatory agency. Alternative design methods such as Bayesian may also be considered if Type I error can be properly controlled.

No matter which design method is used for the basket trial, rejection of the global null hypothesis at the pooled analysis does not mean that the drug is equally effective in all the tumor indications in the pool, or they should be all approved. Similar to that for a fixed sample size trial in terms of impact of baseline characteristics on treatment effect, regulatory decision on drug approval or the scope of the label based on a confirmatory basket trial hinges upon the outcome of additional analyses (e.g., whether the treatment effect in the pooled analysis is driven by a subset of tumor indications, whether the benefit-risk profile of the experimental drug in an individual tumor cohort is favorable). Post marketing studies may be requested to further confirm the clinical benefit.

Complementary to basket trials, an umbrella trial simultaneously investigates multiple experimental drugs in the same tumor indication. The experimental drugs may be added to or removed from an umbrella trial on a rolling basis. It should be randomized whenever multiple experimental arms (or drug cohorts) are unblinded for enrollment. Randomization ratio

may be adapted to emerging data from the trial in favor of the more promising treatment arms, and non-performing arms may be terminated early. Because the drugs are investigated on the same platform, often in a few dedicated sites, there may be less heterogeneity in patient population across the drug cohorts so that the comparison among experimental drugs can be more informative than studied separately.

The randomized controlled umbrella/platform trial is a special type of multi-arm Phase 3 trials thus may follow the same principles of multiplicity adjustment. If the trial is focused on addressing the efficacy questions for each treatment separately, not for a single claim of overall effectiveness, the familywise error rate in a single umbrella/platform trial sharing the common control is always lower than separate trials, and in principle multiplicity adjustment may not be necessary. However, when multiple doses for the same treatment are included in the trial, multiplicity adjustment is required to address the efficacy question for the treatment. Multiplicity control can be substantially complicated with response adaptive randomization or other adaptive features. Primary comparison between the experimental arm and the control arm in a randomized controlled umbrella/platform trial should be generally based on study participants randomized during the same period.

## REFERENCES

- [1] Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures[J]. *Statistics in medicine*, 2009, 28(4): 586-604.
- [2] Chapman P B, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation[J]. *New England Journal of Medicine*, 2011, 364(26): 2507-2516.
- [3] Chen C, Li X N, Li W, et al. Adaptive expansion of biomarker populations in phase 3 clinical trials[J]. *Contemporary clinical trials*, 2018, 71: 181-185.
- [4] Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients[J]. *Clinical cancer research*, 2005, 11(21): 7872-7878.
- [5] Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial[J]. *Clinical trials*, 2006, 3(1): 57-71.
- [6] Hobbs B P, Barata P C, Kanjanapan Y, et al. Seamless designs: current practice and considerations for early-phase drug development in oncology[J]. *JNCI: Journal of the National Cancer Institute*, 2019, 111(2): 118-128.
- [7] Howard D R, Brown J M, Todd S, et al. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group[J]. *Statistical methods in medical research*, 2018, 27(5): 1513-1530.
- [8] ICH. E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2019.
- [9] Ji Y, Liu P, Li Y, et al. A modified toxicity probability interval method for dose-finding trials[J]. *Clinical trials*, 2010, 7(6): 653-663.
- [10] Kang S P, Gergich K, Lubiniecki G M, et al. Pembrolizumab KEYNOTE-001: an adaptive study leading to accelerated approval for two indications and a companion diagnostic[J]. *Annals of oncology*, 2017, 28(6): 1388-1398.
- [11] Mandrekar S J, Sargent D J. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges[J]. *Journal of clinical oncology*, 2009, 27(24): 4027.
- [12] Mayawala K, Tse A, Rubin E H, et al. Dose finding versus speed in seamless immune-oncology drug development[J]. *The journal of clinical pharmacology*, 2017, 57: S143-S145.

- [13] Proschan M A, Follmann D A. Multiple comparisons with control in a single experiment versus separate experiments: why do we feel differently?[J]. The american statistician, 1995, 49(2): 144-149.
- [14] Schwartz L H, Litière S, de Vries E, et al. RECIST 1.1—update and clarification: from the RECIST committee[J]. European journal of cancer, 2016, 62: 132-137.
- [15] Seymour L, Bogaerts J, Perrone A, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics[J]. The lancet oncology, 2017, 18(3): e143-e152.
- [16] Sun L Z, Kang S P, Chen C. Testing monotherapy and combination therapy in one trial with biomarker consideration[J]. Contemporary clinical trials, 2019, 82: 53-59.
- [17] Yuan Y, Hess K R, Hilsenbeck S G, et al. Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials[J]. Clinical cancer research, 2016, 22(17): 4291-4301.
- [18] Zhou H, Yuan Y, Nie L. Accuracy, safety, and reliability of novel phase I trial design[J]. Clinical cancer research, 2018, 24(18): 4357-4364.





