Georgia State University

ScholarWorks @ Georgia State University

Mathematics and Statistics Faculty Publications   Department of Mathematics and Statistics

2012

Methodology and Application of Adaptive and Sequential Approaches

Zhengjia Chen

Yichuan Zhao
Georgia State University, yichuan@gsu.edu

Ye Cui

Jeanne Kowalski

Follow this and additional works at: https://scholarworks.gsu.edu/math_facpub

Part of the Mathematics Commons

Review Article

Methodology and Application of Adaptive and Sequential Approaches in Contemporary Clinical Trials

Zhengjia Chen,[1, 2] Yichuan Zhao,[3] Ye Cui,[3] and Jeanne Kowalski[1]

[1] Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA
[2] Winship Cancer Institute, Emory University, 1365-B Clifton Road, Room B4109, Atlanta, GA 30322, USA
[3] Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA

Correspondence should be addressed to Zhengjia Chen, zchen38@emory.edu

The clinical trial, a prospective study to evaluate the effect of interventions in humans under prespecified conditions, is a standard and integral part of modern medicine. Many adaptive and sequential approaches have been proposed for use in clinical trials to allow adaptations or modifications to aspects of a trial after its initiation without undermining the validity and integrity of the trial. The application of adaptive and sequential methods in clinical trials has significantly improved the flexibility, efficiency, therapeutic effect, and validity of trials. To further advance the performance of clinical trials and convey the progress of research on adaptive and sequential methods in clinical trial design, we review significant research that has explored novel adaptive and sequential approaches and their applications in Phase I, II, and III clinical trials and discuss future directions in this field of research.

1. Clinical Trials

Medicine is of paramount importance for human healthcare. Development of novel successful medicines is a lengthy, difficult, and expensive process which consists of laboratory experimentation, animal studies, clinical trials (Phase I, II, and III), and postmarket followup (Phase IV). Clinical trials are FDA-approved studies conducted in human beings to demonstrate the safety and efficacy of new drugs for health interventions under prespecified conditions. A clinical trial is conducted in a sampled small population and the conclusions reached will be applied to a whole target population; therefore, statistics is an indispensable and critical component of clinical trial development and analysis, which

has become increasingly important in contemporary clinical trials. As the gold standard for the evaluation of a new drug, every contemporary clinical trial must be well designed according to its specific purpose and conducted properly under governmental regulations. The major roles of a statistician in a clinical trial are to design an efficient trial with minimum cost and length and maximum therapeutic effect for patients in the trial, and to draw convincing conclusions by applying appropriate cutting edge statistical knowledge. In the past several decades, numerous groundbreaking novel statistical methodologies have been developed and applied to clinical trials and have significantly improved their performance. Consequently, clinical trials have evolved from simple observation studies to hypothesis-driven and well-designed prospective studies. At present, contemporary clinical trials have become the most important part of modern medicine.

2. Adaptive and Sequential Methods

Classical clinical trials are usually designed with a fixed sample size and schedule without using the information obtained from the ongoing trial. However, it has become increasingly common to modify a trial and/or statistical procedures during the conduct of a clinical trial. Specific modifiable procedures include the patient eligibility and evaluation criteria, drug or treatment dosage and schedule, laboratory testing or clinical diagnosis, study endpoints, measurement of clinical response, formulation of study objectives into statistical hypotheses, appropriate study design according to study purpose, calculation of minimum sample size, participant randomization, study monitoring with interim/futility analysis, statistical data analysis plan, and reaching conclusions, and so forth. The purpose of the modification is to improve the performance of a trial with prompt utilization of data accumulating from within the trial as well as upcoming related information from the literature.

Recently, adaptive and sequential clinical trials have become increasingly popular. The sequential method is an approach of frequentist statistics in which data are evaluated sequentially as they are accumulated and a study is monitored sequentially for stopping whenever a conclusion is reached with enough evidence. Adaptive design refers to the modification of aspects of the trial according to data accumulating during the progress of the trial, while preserving the integrity and validity of the trial. The modifiable aspects of adaptive trials include, but are not limited to, (a) sample size, (b) addition or removal of a study arm, (c) dose modification, (d) treatment switch, and so forth [1]. There are two types of adaptive methods in clinical trials, Bayesian and frequentist approaches [2]. The frequentist approach performs the modification of trials while controlling for type I and II errors. The Bayesian approach allows adaption according to the predicted probability. Common characteristics of sequential and adaptive clinical trials are that the trial and/or statistical procedures are modified during the conduct of trial according to the data accumulating during the trial. The sequential method mainly refers to sequentially monitoring the stopping criteria for futility and efficacy, while adaptive methods include modification of many more aspects of the trial as listed above, in addition to the decision of whether to stop the ongoing trial. Considerable novel statistical research has been conducted in the development of sequential and adaptive methods, especially for Phase I and II clinical trials. However, only some of these methods have actually been applied to the daily practice of real clinical trials. In the next 3 sections, we will review significant sequential and adaptive methods that have been applied to Phase I, II, and III clinical trials and have had a high impact on the field of clinical trials.

3. Statistical Methodology of Phase I Clinical Trials

A Phase I trial is one of the most important steps in a drug's development and is the first clinical trial in human subjects after laboratory and animal studies of a therapeutic agent have shown a potential cure effect on the disease. The sample size of a Phase I clinical trial is relatively small and varies in the range of twenty to eighty. It is a widely accepted assumption that the therapeutic effect of a drug depends on its toxicity and increases monotonically with its dosage level. Higher doses are correlated with both severe toxicity and better therapeutic effect. Therefore, a balance is to be achieved between toxicity level and therapeutic benefit. To achieve the best therapeutic benefit, a patient should be treated with the maximum dosage of drug at which the patient can tolerate its associated toxicities with close monitoring. Among all toxicities patients experience, some are so severe that they limit dose escalation. These toxicities are called dose limiting toxicity (DLT). In the National Cancer Institute (NCI) Common Toxicity Criteria, DLT is defined as a group of grade 3 or higher nonhematologic toxicities and grade 4 hematologic nontransient toxicities. The grades of all toxicities are classified as below:

grade 0: no toxicity;

grade 1: mild toxicity;

grade 2: moderate toxicity;

grade 3: severe toxicity;

grade 4: life-threatening toxicity;

grade 5: death.

The main goals of a Phase I trial are to determine the dose-toxicity relationship of a new therapeutic agent and estimate the maximum tolerated dose (MTD) of the agent given the specified tolerable toxicity level. The highest acceptable DLT level is usually defined as a target toxicity level (TTL). It can be said that the TTL determines the MTD of the new therapeutic agent. A careful and thoughtful approach to the design of Phase I trials and accurate MTD estimation are essential for the fate of the new drug in subsequent clinical trials.

In a Phase I clinical trial, the well accepted assumption is that the probability of toxicity increases monotonically with increasing drug dose, although a decrease in the probability of toxicity at high dose levels could happen in some special cases which are not common and not considered here. There are nonparametric and parametric manners to describe the toxicity-dose relationship. In the non-parametric way, the only assumption is that toxicity is nondecreasing with dose. In the parametric description, a distribution with some parameters is adapted to model the toxicity-dose curve. From a biological point of view, the human body has stabilization and self-salvage systems to protect the person from mild toxicity when a drug dose is at a low level below a certain threshold level, but the probability of toxicity increases at an accelerated speed once the stabilization and self-salvage systems have been overcome, and reaches rapidly the worst condition, death, and then levels off. Therefore a sigmoid shape distribution is an appropriate model to describe the relationship between toxicity probability and dose. Many statistical designs have been proposed for Phase I clinical trials; the most commonly used are summarized and compared in Table 1. According to their algorithm, Phase I clinical trial designs can be grouped into two major categories, rule based design and model based design [3].

Table 1: Summary of main Phase I clinical trial designs.

Designs	Advantages	Disadvantages
Standard 3 + 3 design	Robust. Simple. Easy to carry out.	MTD is not a dose with any particular probability of DLT, but in the range from 20% to 25% DLT. Can not estimate MTD with target probability of DLT <20% or >33%. Not all toxicity data of all patients are used to determine the MTD. Many patients are likely to be treated at low doses.
ID isotonic design	Only assumes a monotonically increasing relationship between dose and toxicity. Semiparametric. Can estimate MTD with different TTL (0~100%). Robust and easy to use. Good for combination of multiple drugs and treatments.	The accuracy of MTD may not be as good as CRM or EWOC. The trial efficiency may not be as good as CRM or EWOC.
CRM continual reassessment method	Fit parametric model for dose toxicity relationship. Adaptive optimal design. Accurate estimation of MTD. Improved trial efficiency. Allow flexible MTD with different TTL.	High risk of patients being treated with over toxic dosages. If the parametric model is not reliable, the result could be questionable. May fail to find MTD.
EWOC escalation with overdose control	Includes all advantages of CRM. Controls the overdosing probability. Further improves MTD accuracy and trial efficiency.	If the parametric model is not reliable, the result could be questionable. May fail to find MTD.

3.1. Rule Based Phase I Designs

All rule based designs follow a sequential approach. In rule based designs, a non-decreasing dose toxicity relationship is the only well accepted assumption required. Therefore rule based designs are well suited for first in human clinical trials in which the dose toxicity relationship is not well understood. Common rule based designs include 3 + 3 design [4], isotonic design [5], accelerated titration design [6], and so forth.

The 3 + 3 designs are rule based up-and-down methods used in Phase I protocol templates of the cancer therapy evaluation program (CTEP), whose mission is to improve the lives of cancer patients by sponsoring clinical trials to evaluate new anticancer agents, with a particular emphasis on translational research to elucidate molecular targets and mechanisms of drug effects. While 3 + 3 designs have become standard practice among many Phase I clinical trialists, they are not designed with the intention of producing accurate estimates of a target quantile. Rather they are designed to screen drugs quickly and identify a dose level that does not exhibit too much toxicity in a very small group of patients. These 3 + 3 designs fall into two categories, without dose de-escalation (Figure 1) and with dose de-escalation (Figure 2). In the 3 + 3 design without dose de-escalation, three patients are assigned to the first dose level. If no DLT is observed, the trial proceeds to the next dose level and another cohort of three patients is enrolled. If at least two out of the three patients experience at least one DLT, then the previous dose level is considered as the MTD; otherwise, if only one patient experiences the DLT, then three additional patients are enrolled at the same dose level.
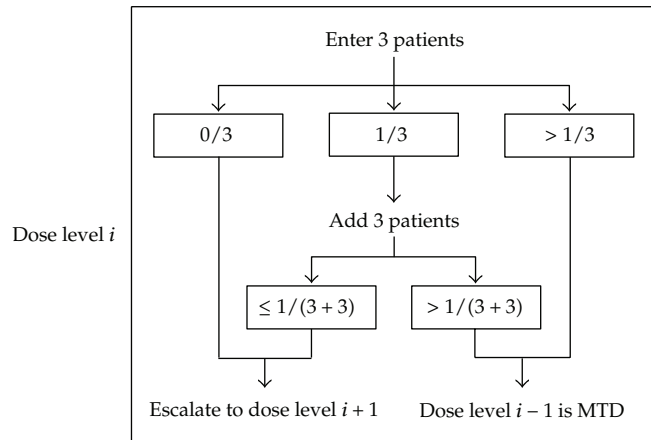
Figure 1: Escalation scheme for 3 + 3 design without dose deescalation (adapted from Lin and Shih [4]).
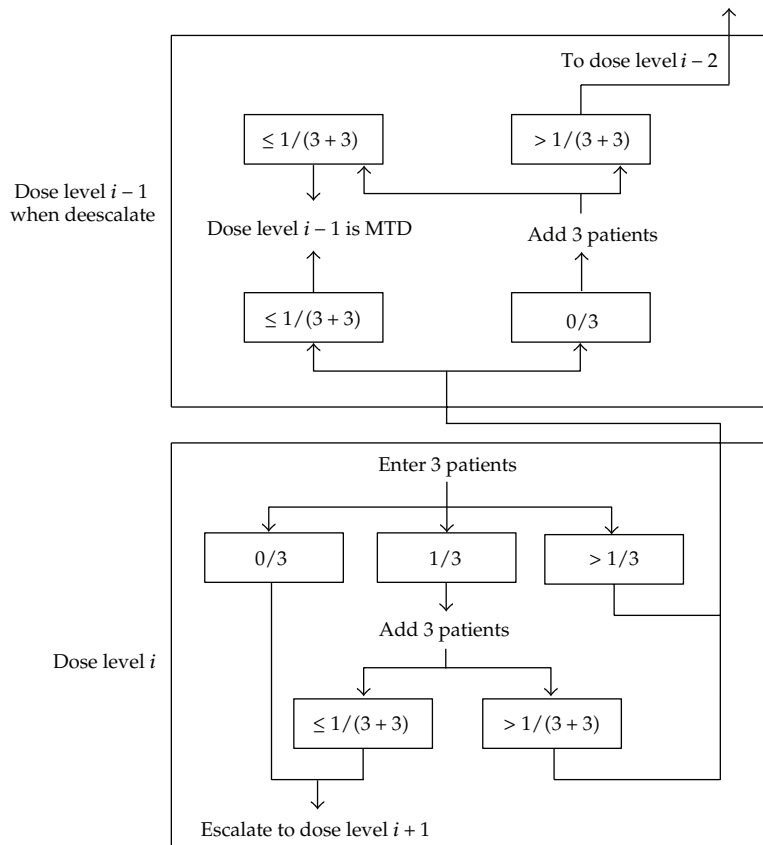

Figure 2: Escalation scheme for 3 + 3 design with dose de-escalation (adapted from Lin and Shih [4]).

If at least one of the three additional patients experiences the DLT, then the previous dose is considered as the MTD; otherwise, the dose will be escalated. The 3 + 3 design with dose de-escalation allows three new patients to be treated at a previous dose level if only three patients were treated at that level previously. Dose reduction continues until a dose level is reached at which six patients are treated and at most one DLT is observed in the six patients. The MTD is defined as the highest dose level at which at most one of six patients experiences DLT, and the immediate higher dose level has at least two patients who experience DLTs. If the first dose is not tolerable, then the MTD cannot be established within the confines of the study. Hence, the MTD is identified from the data and is a statistic rather than a parameter. Storer (1989) was probably the first to examine the characteristics of the 3 + 3 design from the standpoint of the statistician [7]. The operating characteristics of the 3 + 3 design were discussed in Lin and Shih (2001) [4]. Note that any design with sampling that is asymmetric about the MTD will yield a biased result; thus the standard design, and all other designs that approach the MTD from below, will tend to yield a low estimate of the MTD. The 3 + 3 designs are simple and can usually determine a reasonable MTD and are thus the most widely used methods for Phase I clinical trials. But they also have many shortcomings; for example, the methods are not designed around a quantile of interest; not all toxicity data are used to determine the MTD; the MTD is not a dose with any particular probability of toxicity. These disadvantages led to the exploration of extended isotonic design for Phase I clinical trials.

Leung and Wang (2001), for the first time, introduced a semiparametric Phase I design called isotonic design in which only a non-decreasing dose toxicity relationship is the required assumption [5]. In their isotonic design, the pool-adjacent-violators algorithm (PAVA) and isotonic regression are used to update the probability of DLT of each dose level after the toxicity response of each newly treated cohort has been obtained. The dose allocation rationale is to treat each new cohort at a dose level with an estimated probability of DLT closer to the pre-specified target acceptable toxicity level. The trial stops when the same dose has been tested consecutively for a certain number of cohorts or a maximum number of patients have been treated. The recommended dose level for the next cohort based on all completed data after the trial stops is the MTD. Through simulation studies, the isotonic design was demonstrated to perform substantially better than the 3 + 3 design and comparably to the continual reassessment method (CRM) [8], Storer's up-and-down designs, and escalation with overdose control (EWOC) design [9]. Moreover, the isotonic design is model-free and especially appropriate in cases where the parametric dosetoxicity relationship is not well understood.

There are many other rule based designs. All rule based designs can estimate a reasonable MTD using a stopping rule based either on observed DLTs or on convergence criteria. Ad hoc additional dose levels can also be added when needed without any impact on their robustness. Most rule-based designs are practically simple and easy to implement. At present, 3 + 3 designs are still the most popular in Phase I clinical trials.

3.2. Model Based Designs

In model based designs, three parametric dose-toxicity functions (logistic model, hyperbolic model, and power function) are usually employed to depict the relationship between dose and toxicity. Model based designs often fail to find an MTD in first in human studies that are based on observed DLTs. The most common model based designs are CRM and EWOC. Their algorithms are illustrated in Figure 3.
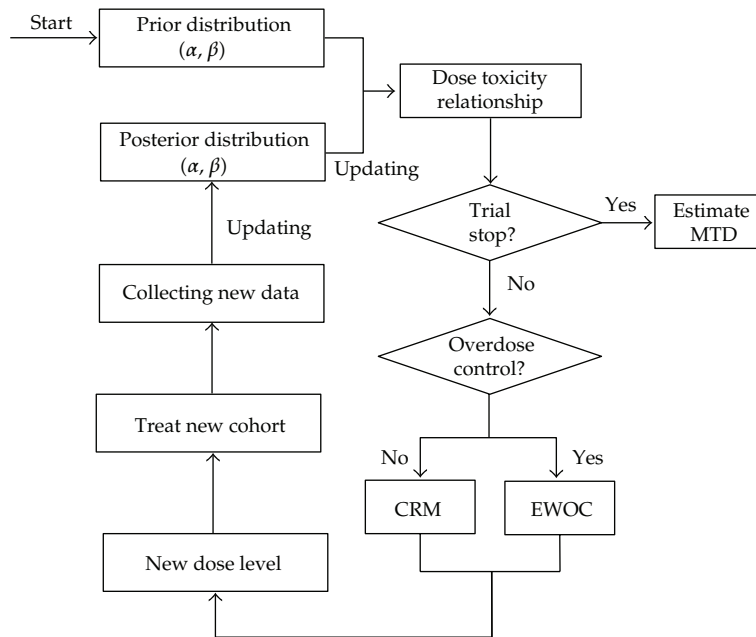
Figure 3: Diagram of model based phase I designs: continuous reassessment design (CRM) and escalation with overdose control (EWOC).

O'Quigley et al. (1990) originally introduced the CRM, a Bayesian approach to fully and efficiently use all data and prior information available in a Phase I study [8]. As in rule based designs, a TTL is specified and the goal is to estimate the dose associated with the TTL, Γ. A parametric model depicting the dose toxicity relationship and a *prior* distribution for each unknown parameter of the model are required to implement CRM. The posterior mean of each parameter is computed using the prior for the parameter and all available toxicity data for the probability of toxicity, P_{DLT}, of each dose level. The computation is conducted and P_{DLT} of each dose level is updated with accumulative toxicity data available when a new patient is recruited. The main idea of CRM is to treat each patient at the dose level with P_{DLT} closest to Γ. The MTD is defined as the dose level of the last patient treated in the trial. In the originally proposed CRM, a one parameter model of dose toxicity function and a single patient cohort are used. Furthermore, the first patient is proposed to be treated at a dose level determined purely by a guess in the original CRM, which makes the method impractical. Therefore, Korn et al. (1994) proposed a modified CRM in which the trial starts at the lowest dose level, no dose level can be skipped during the dose escalation, and the trial stops when the same dose has been recommended for a new patient consecutively for a fixed number of times [10]. However, patients still may be treated at excessively toxic doses in the modified CRM because of its single patient per cohort and the length of study is still very long because of the restriction that the toxicity of all treated patients must be obtained to calculate the new dose level for the next patient. In addition to the modification of Korn at al. (1994) [10], Faries (1994) [11], in his modified CRM, added another rule that no dose escalation is allowed for the next patient when the last patient has DLT. This rule can avoid treating patients at overly toxic doses compared with the traditional 3 + 3 design. In order to address the ethical requirement that the probability of a patient being treated at overdose

is under a pre-specified value, Babb et al. (1998) introduced an adaptive dose escalation scheme called EWOC [9]. The constraint on overdosing of EWOC is a superior feature over the CRM and its theoretical foundation was further elaborated by Zacks et al. (1998) [12]. A two-parameter model logit $(P_{\mathrm{DLT}}(x_i)) = \alpha + \beta x_i$ was first used to depict the dose, x_i, and DLT relationship and then the joint posterior for α and β was transformed to a joint posterior for the MTD and the probability of DLT at the lowest dose level, ρ_0. EWOC is also designed to rapidly approach the MTD in addition to the overdose constraint so that it starts from the lowest dose level and a single patient per cohort is used. After the toxicity response of the last enrolled patient has been obtained, the joint posterior for the MTD and ρ_0 is updated using all the available information and the next coming patient is treated at the 25th percentile of the marginal posterior for the MTD. The trial stops after a fixed number of patients have been treated and then the MTD is computed as its posterior mean or estimated by minimizing the posterior expected loss in a loss function. In order to be safe and shorten the length of the trial, no dose level can be skipped during the dose escalation procedure and multiple patient cohorts can be used instead in EWOC. Through simulation studies, EWOC has been shown to be effective in overdose control and have comparable accuracy of estimated MTD as CRM. Fewer patients are treated at nonoptimal dose levels, resulting in less DLT, and the estimated MTD has smaller average bias and mean squared error in EWOC than in some other nonparametric designs, such as four up-and-down designs and two stochastic approximation methods [9]. It seems that EWOC is a promising alternative design for Phase I clinical trials, especially when the ethical and safety requirement of overdose control is a particular concern. Both CRM and EWOC belong to adaptive dose finding designs in which a Bayesian approach is usually employed and the dose level for the new incoming cohort is adaptive based on the toxicity responses of the previously treated patients in the ongoing trial. Another adaptive dose design is the nonparametric adaptive urn design approach for estimating a dose-response curve [13].

All ruled based designs are robust and simple to implement and usually give a reasonable MTD under certain rules. Applying some sort of models, such as isotonic regression, to data can improve the accuracy of the MTD. Model based designs require a parametric model of dose toxicity relationship and may greatly improve the probability of estimating the correct MTD compared with rule based designs when certain assumptions are satisfied. However, model based designs are not robust and should not be used unless their underlying assumptions can be met with confidence. The accuracy of the estimated MTD depends substantially on the number of observed DLTs, and the sample size is also an important factor. Overall, different designs, whether rule based or model based, usually perform similarly when they are similar in sample size and aggressiveness. Thus, simple designs, especially standard designs, are still very popular in Phase I clinical trial practices.

The design of Phase I clinical trials can involve one or two stages. Rule based or model based designs can be implemented in each stage of two stage designs. There are other critical issues in Phase I clinical trial designs, such as the operating characteristics of 3 + 3 design in terms of expected toxicity level [14], two or multiple stage Phase I design, within-patient dose escalation, late toxicity, combination of multiple agents, balance between toxicity and efficacy, individual MTD, fully utilization of all toxicities [15, 16], and so forth. Some outstanding research studies have been conducted on these topics, which will not be elaborated on herein due to space constraints but have been described in several comprehensive review articles [3, 17–19].

4. Statistical Methodology of Phase II Clinical Trials

After the safety and MTD of an experimental drug have been established in a Phase I clinical trial, the drug will enter Phase II clinical trials, which initially evaluate the drug's therapeutic effects at the recommended MTD. Phase II trials are sometimes further classified as Phase IIa and IIb studies. Phase IIa trials screen the promising novel experimental agent for significant antidisease activity and Phase IIb trials focus on the drug's improved therapeutic effectiveness over the standard treatment. Phase II studies provide critical information to decide whether further testing of the experimental drug in a large confirmatory Phase III trial is warranted. The surrogate endpoint used in Phase II clinical trials needs to be obtained in a short time and should be able to assess the treatment's primary benefit. For cancer trials, the experimental drug's antitumor activity and progression-free survival (PFS) of treated patients are often used as surrogates of the drug's efficacy. The drug's anti-tumor activity is measured as clinical response within a short period of time following the treatment and is classified as complete response (CR), partial response (PR), progressive disease (PD), or stable disease (SD). PFS, which is estimated as the time elapsed from the date of treatment to the date of adverse event, resembles the outcome (overall survival) of the following Phase III clinical trial and is also widely used when it can be measured in a short time.

4.1. Single Arm Phase II Designs

The most commonly used Phase II clinical trial designs are summarized in Table 2. Phase II trials can involve either a single arm, which compares the new treatment with the standard response rate reported by historical data, or two or more arms with patients randomized among different treatments. In a single arm Phase II trial, two or multistage designs may be used to improve the trial efficiency and save resources with early termination of a futile trial. The interim analysis between the consecutive stages examines the accumulated data and decides whether the trial should stop as suggested by the early evidence of futility or should continue to next stage. The earliest two stage Phase II design was proposed by Gehan et al. in 1961 [20], in which a trial is terminated for futility when no patients enrolled in the first stage show any response or continues with the second stage, enrolling an additional number of patients to estimate a more accurate response rate with additional patient data. This design provides interim monitoring and can rule out ineffective drug with minimized sample size. This design is only appropriate for binary outcomes, which differ from the overall survival endpoint used in the following Phase III trial. Moreover, this design has no statistical testing on agents showing some promise and is not optimized. Therefore, Simon (1989) proposed an optimized two stage Phase II design by controlling both type I and type II errors as well as optimizing the sample sizes in both stages [21]. This design can quickly screen out agents without effectiveness while testing further agents with some promise. The design has two subtypes, optimal and minimax. The optimal subtype minimizes the expected overall sample size with the probability of the trial stopping after only the first stage so that it is appropriate for experimental drugs with a high probability of failure after the first stage. The minimax subtype minimizes the maximum possible sample size when the trial stops after completion of two stages so that it is better for highly promising experimental drugs. As with Gehan's design, Simon's two stage designs are only appropriate for binary outcomes. Other investigators have further proposed to conduct multiple interim analyses in Phase II clinical trials by using multistages. For example, Fleming (1982) [22] and Chang et al. (1987)

Table 2: Summary of main Phase II clinical trial designs.

Designs	Advantages	Disadvantages
One stage one arm design	Compare with historical control. Smallest sample size. Simple.	Delay the evaluation of effectiveness. Historical control may not be valid. Subject to population differences, time trends, evaluation bias, and so forth.
Gehan's two stage design	With interim monitoring. Rule out ineffective drug with minimized sample size.	No testing on agents showing some promise. Only suitable for binary outcome. The endpoint is different from that in following Phase III trial.
Simon's two stage design	The samples in two stages are optimized. Quickly screen out agents without effectiveness while testing further agents with some promise. Two choices: optimal versus minimax	Only suitable for binary outcome. The endpoint is different from that in the following Phase III trial.
Bayesian Phase II design	Flexible monitoring schedule. More efficient and robust.	Intensive computation. Relies heavily on statistician during trial.
Randomized Phase II design	Use of randomization. Reliable control and less bias. More similar to Phase III trial.	Sample size increases. Length of trial increases. Cost increases.
Phase II pick the winner design	Efficient and effective way of comparing two or multiple experimental regimens. Each experimental regimen compared with historical controls.	Not appropriate for comparison of adding an experimental agent to standard regimen.
Phase II screening design	Limits the sample size required for a randomized Phase II comparison. Good for comparison of the addition of an experimental agent to standard regimen.	No statistical comparison between the selected arms.
Phase II randomized discontinuation design	Good when significant continued benefit after initial benefit implies significant benefit overall, and vice versa, or when benefit is restricted to a nonidentifiable subgroup of patients.	May need a large number of patients treated at a treatment not effective for them.
Phase II/III design	Use of Phase II data in Phase III trial. Minimize delay in starting up Phase III study. Use of concurrent control. Useful for new drugs showing efficacy.	Large sample sizes. Needs Phase III infrastructure developed even if it stops early.

[23] studied multiple testing and group sequential methods for Phase II trial designs. But the issue of inflating overall type I error needs to be considered in these kinds of Phase II designs.

Among the single arm Phase II designs, another major group is Bayesian Phase II design. For example, Thall and Simon (1994) [24] proposed a Bayesian Phase II design which continuously examines the results after each new enrolled patient and determines whether the trial can stop with a solid decision on the efficacy of the experimental drug or should continue to enroll more patients and obtain enough data for making a decision. Lee and Liu (2008) [25] proposed a Bayesian approach called predictive probability Phase II design. This novel Bayesian design provides a flexible monitoring schedule for Phase II clinical trials which becomes more efficient and robust, but at the cost of intensive computation, and relies heavily on the statistician during the trial. Yin et al. (2011) further coupled the methods of

predictive probability monitoring and adaptive randomization in a randomized Phase II trial and extensively compared this hybrid Bayesian approach with group sequential methods [26].

4.2. Two or More Arm Phase II Designs

Some Phase II clinical trials may have two arms and randomization is frequently used to generate a reliable concurrent control arm and reduce biases. This kind of randomized Phase II trial is more similar to a Phase III trial. Randomized Phase II trials may reduce the so-called trial effect which often arises due to different patient populations, physician preferences, and medical environments between current and previous studies. But the sample size, trial length, and cost increase about 4-fold.

There are several multiple arm Phase II designs [27]. The Phase II "pick the winner" design is one in which each experimental regimen is compared with a historical control. No formal statistical comparison between groups is conducted and the simple winner of the all arms is the winner of the trial. This design provides an efficient and effective way of comparing two or multiple experimental regimens but is not appropriate for the comparison of adding an experimental agent to a standard regimen.

Phase II screening design is another Phase II design with multiple arms in which all experimental arms are compared with the standard treatment arm and all the experimental arms beating the standard treatment arm are winners. Therefore this design limits the sample size required for a randomized Phase II comparison and it is appropriate for testing the effect of adding an experimental agent to a standard regimen. However, it provides no statistical comparison between the selected (winning) arms.

Some investigators have proposed a novel Phase II randomized discontinuation design in which all patients receive the same treatment for a period of time and those with stable disease are randomized to continue or discontinue. This design is particularly appropriate when the treatment is known to have better therapeutic effects and it is ethical for all participants to benefit from it, or when the potential subgroup of patients who can benefit from the treatment is unknown before receiving it. However, this design requires a large number of patients to be treated with a treatment not effective for them. Therefore this design has specific applications but is not widely used.

Conventionally, Phase II and III trials are conducted separately in a sequential order and only an experimental drug that has successfully passed a Phase II trial can enter a Phase III trial. The resulting gap between trials and time lag may be unnecessary under certain circumstances. Therefore, a seamless Phase II/III design has been proposed, which uses Phase II data in a Phase III trial and minimizes delay in starting up the Phase III study [28, 29]. Usually the Phase II part is a randomized Phase II trial which uses a concurrent control. This nonstop Phase II/III design is particularly useful for new drugs showing efficacy. It usually requires large sample sizes and requires a Phase III infrastructure to be developed even if it stops early.

4.3. Other Advanced Topics in Phase II Designs

Categorical tumor response has been the most common endpoint in the Phase II clinical trial designs. However, from a statistical standpoint, categorizing a continuous tumor change percentage into a categorical tumor response with 4 levels results in a loss of study power by

not fully utilizing all available data. Several publications have studied extensively the direct utilization of continuous tumor shrinkage as the primary endpoint for the measurement of drug efficacy in Phase II clinical trials [30–32]. The success rate of Phase III oncology trials remains very low (e.g., 50–60%) despite the success demonstrated in the preceding Phase II trials [30]. The relationship between tumor response/tumor shrinkage percentage and overall survival as the gold standard for drug efficacy has been revisited [33]. PFS has the advantage of short follow-up time [34] and has been confirmed as the best estimate of overall survival [35] so that PFS is recommended as the primary endpoint over categorical tumor response in Phase II clinical trials when feasible.

5. Statistical Methodology of Phase III Clinical Trials

If an experimental agent exhibits adequate short term therapeutic effects in a Phase II trial, the drug will be moved forward to a Phase III study for confirmative testing of its long term effectiveness. The typical endpoint in a Phase III trial is a time to event measurement, such as progression free survival or overall survival. Phase III trials are large scale in terms of sample size, resources, efforts, and costs. This Phase collects a large amount of data over a long period of followup to evaluate the ultimate therapeutic effect of a new drug. The design of Phase III clinical trials has become a very important research field in order to improve the performance of these critical clinical trials. The most commonly used Phase III clinical trial designs are summarized in Table 3.

5.1. Randomization

The earliest design of Phase III clinical trials is a single arm study design using historical controls from the literature, existing databases, or medical charts. This kind of Phase III design allows ethical consideration and can increase enrollment as patients are assured of receiving new therapy. In addition, trials will have shorter time and lower cost, making this type of trial a good choice for the initial testing of new treatments, or when disease diagnosis is clearly established, prognosis is well known, or the disease is highly fatal. This Phase III design, however, provides no comparison to control group data and is vulnerable to biases because disease and mortality rates have changed over time and literature controls are particularly poor. Phase III trials conducted using this design tend to exaggerate the value of a new treatment. In order to avoid bias and eliminate time trends, a concurrent control but nonrandomized design for Phase III clinical trials was then proposed and implemented. In this design, randomization does not interfere with treatment selection. It is easier to select a group to receive the intervention and select the controls matching key characteristics. Therefore, this design can reduce costs and is relatively simple and easily acceptable to both the investigator and participant. But in this Phase III design, intervention and control groups may not be comparable because of selection bias and incomparable different group populations. It is difficult to prove comparability because it is impractical to have information on all important prognostic factors and to match several factors. The existence of unknown or unmeasured factors in large studies is also uncertain. The afterward covariance analysis is not adequate for offsetting the imbalance between groups.

To eliminate the bias, facilitate masking treatments, and permit the use of statistical theory, randomization has been employed widely in the Phase III clinical trials [36]. There are two major types of randomization approaches, non adaptive versus adaptive.

Table 3: Summary of main Phase III clinical trial designs.

Designs	Advantages	Disadvantages
Historical control (literature and existing databases or medical charts)	Allows ethical consideration. Increase enrollment as patients are assured of receiving new therapy. Shorter time and less cost. Good for initial testing of new treatments, when disease diagnosis is clearly established, prognosis is well known, or disease is highly fatal.	Vulnerable to bias. Disease rate and mortality rate have changed over time. No comparison to control group data. Literature controls particularly poor. Tends to exaggerate the value of a new treatment.
Concurrent control, not randomized	Eliminates time trends. Data of comparable quality. Randomization does not interfere with treatment selection. Easier to select a group to receive the intervention and select the controls matching key characteristics. Reduced cost, relative simplicity, investigator and participant acceptance.	Intervention and control groups may not be comparable because of selection bias and different treatment groups are not comparable. Difficult to prove comparability because of the need for information on all important prognostic factors and matching several factors is impractical. Uncertainty about unknown or unmeasured factors exists even for large studies. Covariance analysis not adequate.
Randomized clinical trials (RCT)	Considered to be "gold standard". Removes potential bias in group allocation. Randomization and concurrent control produce comparable groups. Guarantees the validity of statistical tests and valid comparison. General use.	Subjects may not represent general patient population. Increased sample size and cost. Acceptability of randomization process. Administrative complexity.
Sequential RCT design	Continues to randomize subjects until null hypothesis is either rejected or "accepted." Good for acute response, paired subjects, and continuous testing. Good for one-time dichotomous decisions such as regulatory approval, and so forth.	Multiple testing inflates type I error. Inhibits adaptation due to the requirement of prespecifying all possible study outcomes.
Bayesian RCT design	Dynamic learning adaptive feature. Incorporates external evidence. Add new interventions and drop less effective ones without restarting trial. Improves timeliness and clinical relevance of trial results. Lowest sample size and cost.	May be criticized as too subjective, not well planned, or too complicated.

Simple randomization, block randomization, and stratified randomization belong to the nonadaptive randomization type. The simple randomization is robust against both selection and accidental biases and appropriate for RCTs with over 200 subjects because of the possibility of imbalanced group sizes in small RCTs [37]. Block randomization can guarantee balanced group sizes by pre-specifying the block size and allocation ratio and allocating subjects randomly within each block [33]. Block randomization is often used with "stratified randomization" in small RCTs. There are several adaptive randomization approaches:

adaptive biased coin, covariate adaptive, and response adaptive [33]. The adaptive biased-coin randomization method can reduce the imbalance of group size and is less affected by selection bias than permuted-block randomization by decreasing and increasing the probability of being assigned to an overrepresented group and underrepresented group, respectively. Randomization can be adaptive to covariate in order to produce balanced groups in terms of the sample size of several covariates. The most common covariate adaptive randomization approaches are the Taves's method [38], Pocock and Simon method [39], and Frane's method [40] for both continuous and categorical types. Overall, covariate adaptive randomization can reduce the imbalance further and handle more covariates simultaneously than using the combination of block and stratified randomization [41]. Randomization can be adaptive to response or outcome in order to increase the trial therapeutic effect, taking into account ethical considerations. Response-adaptive randomization can assign more patients to receive better treatment by skewing the probability of assigning new patients to the group showing favorable response as the data of the trial are accumulating while maintaining a certain study power [41]. The most common approaches used for response-adaptive randomization are the urn model, biased coin design, and Bayesian's approach [34]. Each randomization approach has its own merits and limitations. The selection of randomization method depends on the specific study purpose.

5.2. Randomized Controlled Phase III Trials

The statistical approach of randomization removes any potential bias in group allocation. The use of randomization and a concurrent control together produce comparable groups and make conclusions more convincing. The use of feasible blinding minimizes the bias after randomization. At present, the standard form of a Phase III trial is a randomized and placebo-controlled clinical trial (RCT) with double blinds. The control arm may be a placebo or the standard of care. The use of placebo is only acceptable if there is no other better or standard therapy available. Interim monitoring is also often considered for a long term confirmatory RCT. The RCT which guarantees the validity of statistical tests and valid comparisons has been generally used as the "gold standard" for verifying the efficacy of new drugs. However, there are still some limitations in RCTs; for example, subjects may not represent the general patient population; sample size and cost increase substantially; the randomization process may not be widely accepted; the administrative process may be complex; and so forth. According to their statistical algorithm and characteristics, besides the conventional fixed sample Phase III clinical trial in which only one final data analysis is conducted at the end of the study, other RCT designs with additional analyses before final analysis can be divided into two distinct categories: sequential RCT design and Bayesian adaptive RCT design.

5.2.1. Group Sequential RCT Design

The scheme of the group sequential design is summarized in Figure 4. In this design, type I and II errors are explicitly controlled while testing the study hypotheses, and patients continue to be enrolled and randomized until the primary hypothesis has been proved or disproved. To design a Phase III clinical trial with the group sequential method, the total number of stages, the sample size, and stopping criterion at each stage for the null hypothesis testing as well as the usual specifications in a conventional Phase III clinical trial must be pre-specified before the trial starts. At each interim stage, all accumulated data up to the point are
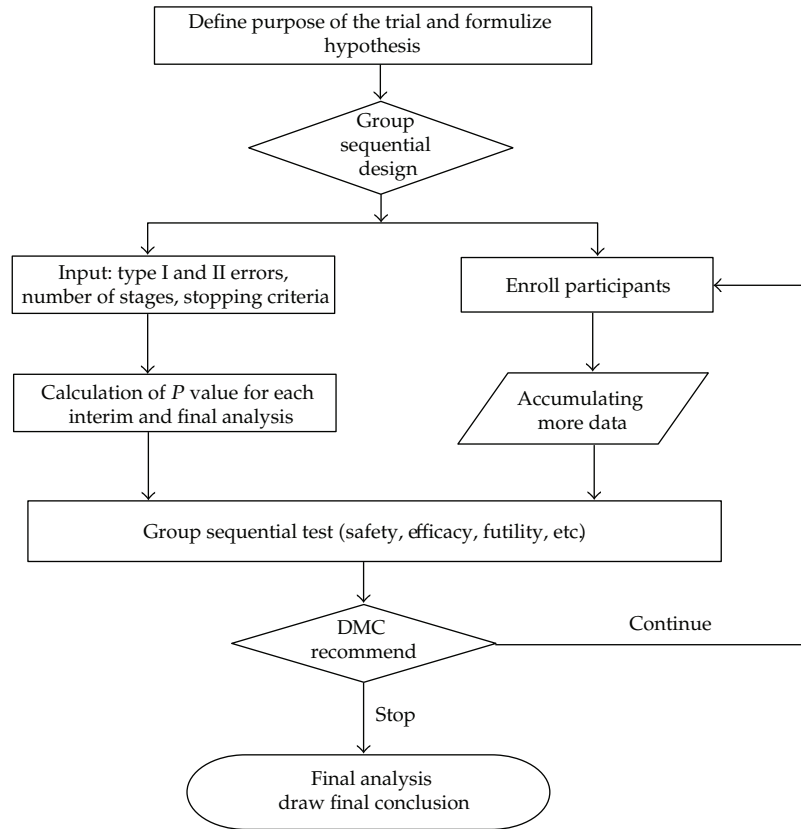
Figure 4: Diagram of group sequential design.

analyzed and the test statistics is compared with critical values generated from the sequential design to determine whether the trial should stop or continue. A conclusion on the primary hypothesis must be reached at the final stage when the sequential trial passes all interim analyses and completes with the final stage.

Multiple testing during the sequential trial may inflate type I error which can be controlled using the Pocock approach [42], O'Brien-Fleming approach [43], and alpha spending function [44]. The Pocock approach was the first method for group sequential testing with given overall type 1 error and power by dividing type I error evenly across the number of interim and final analyses. For example, in a clinical trial with 2 interim analyses and 1 final analysis, the Pocock procedure uses the same cut-off for both the interim and final analyses and the clinical trial can stop and claim a positive outcome if the P value is less than 0.022 at any of the analysis times. One obvious problem with the Pocock approach is its too high probability of stopping the trial early. In order to prevent early stopping and to keep the final P value close to the overall significance level, such as 0.05, O'Brien and Fleming's approach [43] uses a very strict cut-off P value at the beginning, then relaxes the cut-off P value over time. As in the above clinical trial, the P values for the first and second interim analyses are 0.005 and 0.014, respectively. The P value for the final analysis is 0.045 which is close to 0.05. Both the Pocock and O'Brien-Fleming approaches maintain the overall type I error by paying a penalty at the final analysis, but the O'Brien-Fleming

method involves much less of a penalty at the planned conclusion of the study because it requires stricter standards earlier. Both methods have some limitations; both require a pre-specified maximum number of patients, the number of interim analysis, and equal increments of information between interim stages. Therefore, DeMets and Lan [44] (1994) introduced a spending function approach to relax the requirement of the equal increments of information. The approach spends the allowable type I error rate over time according to a chosen spending principle and the amount of information accrued and allows dropping or adding an interim analysis during conduct of the trial. There are several types of spending functions proposed in the literature. Besides the Pocock-type and O'Brien-Fleming-type error spending functions proposed by Lan and DeMets, the gamma error spending function [45] proposed by Hwang, Shih, and DeCani and the power error spending function [46] proposed by Jennison and Turnbull are also commonly used in clinical trials. The conclusions drawn at the interim and final analyses are affected heavily by the pre-specified boundaries so that the choices of the type of spending function are very important and depend on the specific purpose of the trial and its associated clinical program. In addition to efficacy, the safety profile of drug is also an important factor when considering the early stopping of a trial.

The major advantages of the group sequential RCT design are its abilities to prevent unnecessary exposure of patients to an unsafe or ineffective new drug or to a placebo treatment, and to save time and resources by stopping the trial early for efficacy, futility, and safety. The sequential RCT design is suitable for acute response, paired subjects, and continuous testing. It is especially appropriate for dichotomized decisions (yes/no) because the result of the RCT trial is determined to be significant or not according to a pre-specified significance level (type I error). Although sequential RCT is the most widely used design in Phase III clinical trials, it has some limitations. Sequential RCT may require larger sample sizes than Bayesian adaptive RCT as a result of additional variability and comparison of multiple treatments with similar efficacies. Sequential RCT is somewhat adaptive by using interim monitoring and stopping rules, but it requires prespecification of all possible study outcomes, thus inhibiting the full adaptation and utilization of newly accumulated data from the ongoing trial.

5.2.2. Bayesian RCT Design

Bayesian randomized clinical trials refer to trials in which Bayesian approaches are applied extensively to some or all of the processes of a trial including randomization, monitoring, interim and futility analysis, final analysis, and adaptive decisions. Berry and Kadane [47] proposed optimal Bayesian randomization in 1997 and the practical uses of Bayesian adaptive randomization in clinical trials have been reviewed by Thall and Wathen [48]. Bayesian monitoring has been frequently used in some Phase III clinical trials, especially in those with failure time endpoints [49]. Bayesian analysis in clinical trials has become increasingly common recently as it can borrow strength from outside the study [50]. Bayesian adaptive decisions in clinical trials can be made according to a posterior probability or predictive probability of trial success or from the result of Bayesian final analysis. Bayesian adaptive decisions have been compared to frequentist sequential approaches [51] and some studies [52–54] proposed to use Bayesian decision theoretical approaches in the optimization of designs under various settings.

Bayesian RCT design is dynamic learning adaptive in nature as it prespecifies the approaches to combine all available data accumulated during the process of the study,

calculate probabilistic estimation of uncertainty, control the probability of false-positive and false-negative conclusions, and change the study design correspondingly [55]. Bayesian and adaptive RCT design cannot only compare multiple active treatments but can also allow the ongoing trial to add new emerging effective interventions, discontinue less effective ones proved by accumulated within-trial data, or focus on patient subgroups identified by certain biomarkers for whom interventions are more (or less) effective so that the trial tests the most current interventions, improves the clinical relevance, and targets biomarkers that predict response to alternative intervention. Using external existing data from previous studies during the design stage and the accumulated within-trial data to update the design results in smaller sample size, shorter time, and reduced cost of Bayesian and adaptive RCT [56]. But Bayesian RCT may be criticized as being too subjective, not well planned, or too complicated.

Both Bayesian and sequential RCT designs have their advantages and disadvantages. Instead of biasing toward either Bayesian or sequential methods, statisticians and investigators should choose the design of Phase III clinical trial that best fits the goals of the trial and is most likely to provide the best performance.

5.2.3. Adaptive Sample Size Calculation and Adaptive Stopping

In the planning stage of a Phase III clinical trial, sample size is one of the most important factors to be considered because the budget for the trial depends on the minimum required sample size. Usually sample size is fixed in a trial, but an adaptive sample size calculation is often used in adaptive clinical trials and the sample size is adjusted based on the observed data at the interim analysis [1]. Sample size determination depends on the expected treatment difference and its standard deviation; however, their initial estimations often turn out to be too large or small as suggested by the accumulating data from the ongoing trial or other newly completed studies. In this case, keeping the original sample size will lead to an underpowered or overpowered trial, and so the sample size should be adjusted according to the updated effect size for the ongoing trial. There are several approaches for sample size adjustment based on the criteria of treatment effect size, conditional power, and/or reproducibility probability [57–61]. The observed treatment effect and estimated standard deviation from a limited number of subjects at the interim analysis may not be of statistical significance. Therefore, these factors should not be weighed too heavily and the targeted clinically meaningful difference in the ongoing clinical trial should always be considered fully in the adaptive sample size calculation.

The fate of an ongoing Phase III trial is determined at its data monitoring committee (DMC) meeting, which makes recommendations based on the available data according to stopping rules in the statistical guidelines. The common factors considered in stopping rules are safety, efficacy, futility, benefit-risk ratio, weight between the short term and long term treatment effects, and conditional power or predictive power [1]. Current tools for monitoring Phase III trials are stopping boundaries, conditional and predictive powers, futility index, repeated confidence interval, and Bayesian monitoring tools. Even though the stopping rules are usually stipulated in the design stage, adaptive stopping is becoming more and more common due to unpredicted events during the conduct of the trial, such as a change in the DMC meeting date because of unavailability of committee members, different patient accrual progress, and deviation in the analysis schedule. Moreover, the true variability in the parameters to construct these boundaries of stopping rules is never known and it is very common that the initial estimates of the variability and treatment effect in

the design phase are inaccurate as shown by the preliminary results of the ongoing trials. These deviations could affect substantially the stopping boundaries so that adaptive stopping becomes especially desirable in these cases. To stop a trial prematurely under adaptive stopping algorithm, thresholds for the number of subjects randomized and some rules (such as utility rules, futility rules, etc.) in terms of boundaries must pass.

6. Concluding Remarks

Clinical trials remain an indispensable component of new drug development. Novel statistical approaches have been applied to clinical trials and have significantly improved their performance in every step from design, conduct, and monitoring to data analysis and drawing final conclusions. As modern medicine progresses, increasingly complex requirements and factors need to be considered in clinical trials, which in turn create new challenges for statisticians. In the future, more novel statistical approaches, frequentist and Bayesian, should be developed to enhance the performance of clinical trials in terms of therapeutic effect, safety, accuracy, efficiency, simplicity, and validity of conclusions and to expedite the development of effective new drugs to improve human healthcare.
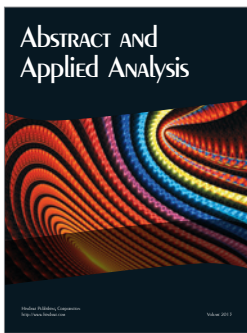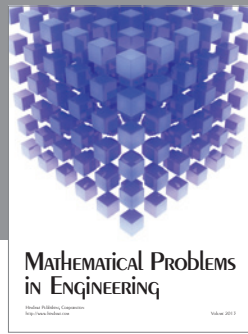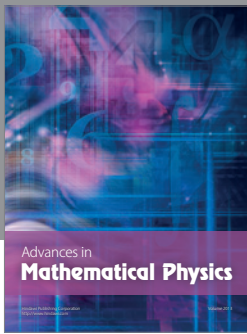
Acknowledgments

References

[1] S. C. Chow and M. Chang, "Adaptive design methods in clinical trials—a review," *Orphanet Journal of Rare Diseases*, vol. 3, no. 1, article 11, 2008.

[2] P. Gallo, C. Chuang-Stein, V. Dragalin, B. Gaydos, M. Krams, and J. Pinheiro, "Adaptive designs in clinical drug development—an executive summary of the PhRMA working group," *Journal of Biopharmaceutical Statistics*, vol. 16, no. 3, pp. 275–283, 2006.

[3] D. M. Potter, "Phase I studies of chemotherapeutic agents in cancer patients: a review of the designs," *Journal of Biopharmaceutical Statistics*, vol. 16, no. 5, pp. 579–604, 2006.

[4] Y. Lin and W. J. Shih, "Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials," *Biostatistics*, vol. 2, no. 2, pp. 203–215, 2001.

[5] D. H. Y. Leung and Y. G. Wang, "Isotonic designs for phase I trials," *Controlled Clinical Trials*, vol. 22, no. 2, pp. 126–138, 2001.

[6] R. Simon, B. Freidlin, L. Rubinstein, S. G. Arbuck, J. Collins, and M. C. Christian, "Accelerated titration designs for phase I clinical trials in oncology," *Journal of the National Cancer Institute*, vol. 89, no. 15, pp. 1138–1147, 1997.

[7] B. E. Storer, "Design and analysis of phase I clinical trials," *Biometrics*, vol. 45, no. 3, pp. 925–937, 1989.

[8] J. O'Quigley, M. Pepe, and L. Fisher, "Continual reassessment method: a practical design for phase 1 clinical trials in cancer," *Biometrics*, vol. 46, no. 1, pp. 33–48, 1990.

[9] J. Babb, A. Rogatko, and S. Zacks, "Cancer phase I clinical trials: efficient dose escalation with overdose control," *Statistics in Medicine*, vol. 17, no. 10, pp. 1103–1120, 1998.

[10] E. L. Korn, D. Midthune, T. Timothy Chen et al., "A comparison of two phase I trial designs," *Statistics in Medicine*, vol. 13, no. 18, pp. 1799–1806, 1994.

[11] D. Faries, "Practical modifications of the continual reassessment method for phase I cancer clinical trials," *Journal of Biopharmaceutical Statistics*, vol. 4, no. 2, pp. 147–164, 1994.

[12] S. Zacks, A. Rogatko, and J. Babb, "Optimal Bayesian-feasible dose escalation for cancer phase I trials," *Statistics & Probability Letters*, vol. 38, no. 3, pp. 215–220, 1998.

[13] R. Mugno, W. Zhus, and W. F. Rosenberger, "Adaptive urn designs for estimating several percentiles of a dose-response curve," *Statistics in Medicine*, vol. 23, no. 13, pp. 2137–2150, 2004.

[14] Z. Chen, M. D. Krailo, J. Sun, and S. P. Azen, "Range and trend of expected toxicity level (ETL) in standard A + B designs: a report from the children's oncology group," *Contemporary Clinical Trials*, vol. 30, no. 2, pp. 123–128, 2009.

[15] Z. Chen, M. D. Krailo, S. P. Azen, and M. Tighiouart, "A novel toxicity scoring system treating toxicity response as a quasi-continuous variable in phase I clinical trials," *Contemporary Clinical Trials*, vol. 31, no. 5, pp. 473–482, 2010.

[16] Z. Chen, M. Tighiouart, and J. Kowalski, "Dose escalation with overdose control using a quasi-continuous toxicity score in cancer phase I clinical trials," *Contemporary Clinical Trials*, vol. 33, no. 5, pp. 949–958, 2012.

[17] M. Tighiouart and A. Rogatko, "Dose finding with escalation with overdose control (EWOC) in cancer clinical trials," *Statistical Science*, vol. 25, no. 2, pp. 217–226, 2010.

[18] W. F. Rosenberger and L. M. Haines, "Competing designs for phase I clinical trials: a review," *Statistics in Medicine*, vol. 21, no. 18, pp. 2757–2770, 2002.

[19] C. Le Tourneau, J. J. Lee, and L. L. Siu, "Dose escalation methods in phase I cancer clinical trials," *Journal of the National Cancer Institute*, vol. 101, no. 10, pp. 708–720, 2009.

[20] E. A. Gehan, "The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent," *Journal of Chronic Diseases*, vol. 13, no. 4, pp. 346–353, 1961.

[21] R. Simon, "Optimal two-stage designs for phase II clinical trials," *Controlled Clinical Trials*, vol. 10, no. 1, pp. 1–10, 1989.

[22] T. R. Fleming, "One-sample multiple testing procedure for phase II clinical trials," *Biometrics*, vol. 38, no. 1, pp. 143–151, 1982.

[23] M. N. Chang, T. M. Therneau, H. S. Wieand, and S. S. Cha, "Design for group sequential phase II clinical trials," *Biometrics*, vol. 43, no. 4, pp. 865–874, 1987.

[24] P. F. Thall and R. Simon, "A Bayesian approach to establishing sample size and monitoring criteria for phase II clinical trials," *Controlled Clinical Trials*, vol. 15, no. 6, pp. 463–481, 1994.

[25] J. J. Lee and D. D. Liu, "A predictive probability design for phase II cancer clinical trials," *Clinical Trials*, vol. 5, no. 2, pp. 93–106, 2008.

[26] G. Yin, N. Chen, and J. J. Lee, "Phase II trial design with Bayesian adaptive randomization and predictive probability," *Journal of the Royal Statistical Society C*, vol. 61, no. 2, pp. 219–235, 2012.

[27] L. Rubinstein, J. Crowley, P. Ivy, M. Leblanc, and D. Sargent, "Randomized phase II designs," *Clinical Cancer Research*, vol. 15, no. 6, pp. 1883–1890, 2009.

[28] P. F. Thall, "A review of phase 2-3 clinical trial designs," *Lifetime Data Analysis*, vol. 14, no. 1, pp. 37–53, 2008.

[29] L. Y. T. Inoue, P. F. Thall, and D. A. Berry, "Seamlessly expanding a randomized phase II trial to phase III," *Biometrics*, vol. 58, no. 4, pp. 823–831, 2002.

[30] P. T. Lavin, "An alternative model for the evaluation of antitumor activity," *Cancer Clinical Trials*, vol. 4, no. 4, pp. 451–457, 1981.

[31] Y. Wang, C. Sung, C. Dartois et al., "Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development," *Clinical Pharmacology and Therapeutics*, vol. 86, no. 2, pp. 167–174, 2009.

[32] T. G. Karrison, M. L. Maitland, W. M. Stadler et al., "Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non-small-cell lung cancer," *Journal of the National Cancer Institute*, vol. 99, no. 19, pp. 1455–1461, 2007.

[33] M. W. An, S. J. Mandrekar, M. E. Branda et al., "Comparison of continuous versus categorical tumor measurement–based metrics to predict overall survival in cancer treatment trials," *Clinical Cancer Research*, vol. 17, no. 20, pp. 6592–6599, 2011.

[34] G. Yothers, "Toward progression-free survival as a primary end point in advanced colorectal cancer," *Journal of Clinical Oncology*, vol. 25, no. 33, pp. 5153–5154, 2007.

[35] M. Buyse, P. Thirion, R. W. Carlson, T. Burzykowski, G. Molenberghs, and P. Piedbois, "Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer," *The Lancet*, vol. 356, no. 9227, pp. 373–378, 2000.

[36] K. F. Schulz and D. A. Grimes, "Generation of allocation sequences in randomised trials: chance, not choice," *The Lancet*, vol. 359, no. 9305, pp. 515–519, 2002.

[37] J. M. Lachin, J. P. Matts, and L. J. Wei, "Randomization in clinical trials: conclusions and recommendations," *Controlled Clinical Trials*, vol. 9, no. 4, pp. 365–374, 1988.

[38] D. R. Taves, "Minimization: a new method of assigning patients to treatment and control groups," *Clinical Pharmacology and Therapeutics*, vol. 15, no. 5, pp. 443–453, 1974.

[39] S. J. Pocock and R. Simon, "Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial," *Biometrics*, vol. 31, no. 1, pp. 103–115, 1975.

[40] J. W. Frane, "A method of biased coin randomization, its implementation, and its validation," *Drug Information Journal*, vol. 32, no. 2, pp. 423–432, 1998.

[41] M. Kang, B. G. Ragan, and J. H. Park, "Issues in outcomes research: an overview of randomization techniques for clinical trials," *Journal of Athletic Training*, vol. 43, no. 2, pp. 215–221, 2008.

[42] S. J. Pocock, "Group sequential methods in the design and analysis of clinical trials," *Biometrika*, vol. 64, no. 2, pp. 191–199, 1977.

[43] P. C. O'Brien and T. R. Fleming, "A multiple testing procedure for clinical trials," *Biometrics*, vol. 35, no. 3, pp. 549–556, 1979.

[44] D. L. DeMets and K. K. G. Lan, "Interim analysis: the alpha spending function approach," *Statistics in Medicine*, vol. 13, no. 13-14, pp. 1341–1352, 1994.

[45] I. K. Hwang, W. J. Shih, and J. S. Decani, "Group sequential designs using a family of type I error probability spending functions," *Statistics in Medicine*, vol. 9, no. 12, pp. 1439–1445, 1990.

[46] C. Jennison and B. W. Turnbull, *Group Sequential Methods with Applications to Clinical Trials*, CRC Press, 2000.

[47] S. M. Berry and J. B. Kadane, "Optimal Bayesian randomization," *Journal of the Royal Statistical Society B*, vol. 59, no. 4, pp. 813–819, 1997.

[48] P. F. Thall and J. K. Wathen, "Practical Bayesian adaptive randomisation in clinical trials," *European Journal of Cancer*, vol. 43, no. 5, pp. 859–866, 2007.

[49] G. L. Rosner, "Bayesian monitoring of clinical trials with failure-time endpoints," *Biometrics*, vol. 61, no. 1, pp. 239–245, 2005.

[50] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, Boca Raton, Fla, USA, 1995.

[51] D. A. Berry and C. H. Ho, "One-sided sequential stopping boundaries for clinical trials: a decision-theoretic approach," *Biometrics*, vol. 44, no. 1, pp. 219–227, 1988.

[52] J. D. Eales and C. Jennison, "An improved method for deriving optimal one-sided group sequential tests," *Biometrika*, vol. 79, no. 1, pp. 13–24, 1992.

[53] N. Cressie and J. Biele, "A sample-size-optimal Bayesian procedure for sequential pharmaceutical trials," *Biometrics*, vol. 50, no. 3, pp. 700–711, 1994.

[54] S. Barber and C. Jennison, "Optimal asymmetric one-sided group sequential tests," *Biometrika*, vol. 89, no. 1, pp. 49–60, 2002.

[55] M. Krams, K. R. Lees, W. Hacke, A. P. Grieve, J. M. Orgogozo, and G. A. Ford, "Acute stroke therapy by inhibition of neutrophils (ASTIN): an adaptive dose-response study of UK-279,276 in acute ischemic stroke," *Stroke*, vol. 34, no. 11, pp. 2543–2548, 2003.

[56] B. R. Luce, J. M. Kramer, S. N. Goodman et al., "Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change," *Annals of Internal Medicine*, vol. 151, no. 3, pp. 206–209, 2009.

[57] M. Posch and P. Bauer, "Adaptive two stage designs and the conditional error function," *Biometrical Journal*, vol. 41, no. 6, pp. 689–696, 1999.

[58] L. Cui, H. M. J. Hung, and S. J. Wang, "Modification of sample size in group sequential clinical trials," *Biometrics*, vol. 55, no. 3, pp. 853–857, 1999.

[59] W. J. Shih, "Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: a comparison," *Statistics in Medicine*, vol. 25, no. 6, pp. 933–941, 2006.

[60] M. A. Proschan and S. A. Hunsberger, "Designed extension of studies based on conditional power," *Biometrics*, vol. 51, no. 4, pp. 1315–1324, 1995.

[61] M. A. Proschan, "Two-stage sample size re-estimation based on a nuisance parameter: a review," *Journal of Biopharmaceutical Statistics*, vol. 15, no. 4, pp. 559–574, 2005.

Advances in
Mathematical Physics

The Scientific
World Journal

Mathematical Problems
in Engineering

Advances in
Decision
Sciences

International Journal of
Combinatorics

Abstract and
Applied Analysis

International Journal of
Stochastic Analysis

Discrete Dynamics
in Nature and Society

Hindawi

Submit your manuscripts at
http://www.hindawi.com

Journal of Function Spaces
and Applications

Journal of
Applied Mathematics

Advances in
Operations
Research

International Journal of
Differential
Equations

Journal of
Probability
and
Statistics

International
Journal of
Mathematics and
Mathematical
Sciences

ISRN
Mathematical
Analysis

ISRN
Discrete
Mathematics

ISRN
Algebra

ISRN
Applied
Mathematics

ISRN
Geometry