HHS Public Access

Author manuscript

Ther Innov Regul Sci. Author manuscript; available in PMC 2015 May 04.

Published in final edited form as:

Ther Innov Regul Sci. 2014 January ; 48(1): 81–89. doi:10.1177/2168479013501309.

Advances in Statistical Approaches Oncology Drug Development

Anastasia Ivanova, PhD [Associate Professor],
Department of Biostatistics, University of North Carolina at Chapel Hill, NC, USA

Gary L. Rosner, ScD [Professor],
Oncology Biostatistics & Bioinformatics, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA

Olga Marchenko, PhD [Vice President],
CSDD, Innovation, Quintiles, Durham, NC, USA

Tom Parke, BSc [Head of Clinical Trial Solutions],
Tessella, Abingdon, Oxfordshire, England

Inna Perevozskaya, PhD [Senior Director], and
Statistical Research and Consulting Center, Pfizer, Inc., Collegeville, PA, USA

Yanping Wang, PhD [Research Advisor]
Biometrics and Advanced Analytics, Eli Lilly and Company, Indianapolis, IN, USA

Abstract

We describe some recent developments in statistical methodology and practice in oncology drug development from an academic and an industry perspective. Many adaptive designs were pioneered in oncology, and oncology is still at the forefront of novel methods to enable better and faster Go/No-Go decision making while controlling the cost.

Keywords

cancer trial; oncology drug development; MTD; model-based design; stopping rule

Introduction

Oncology drug development is quite different from drug development in other therapeutic areas. Phase I oncology trials enroll actual patients rather than healthy volunteers. As a result, investigators and patients are more accepting of toxicity or adverse events, and the nature of toxicity is dose-dependent for most oncology agents. The main objective of phase I is to find the maximum tolerated dose (MTD) as efficiently and safely as possible. Phase I trials are small, with cohorts of patients evaluated at different doses until dose-limiting toxicities (DLTs) are reached. The main objective of phase II is to evaluate the anti-cancer

Correspondence Address: Anastasia Ivanova, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill NC 27599-7420, USA, aivanova@bios.unc.edu.

activity of a drug and provide further safety evaluation. Phase II trials are generally non-randomized one-arm studies in which patients receive the dose determined in previous phase I studies, namely, the MTD in the US and one dose below the MTD in Europe and Japan. Traditionally, phase III has been the first randomized controlled study. Phase III clinical trials are large and complex trials that focused on being the definitive assessment how effective and safe a drug is compared to a current gold standard treatment. One positive phase III trial is generally sufficient for drug approval in most oncology indications. Despite being expensive, the failure rate of phase III trials in oncology drug development is high. There is an unmet need in pharmaceutical industry for better, more efficient designs that could enable better and faster decision making while controlling the cost.

In this paper we describe some recent developments in statistical methodology and practice in oncology drug development from an academic and an industry perspective.

Novel Methods in Phase I Trials in Oncology

The traditional or 3+3 design (1) has been widely used in oncology phase I trials. More efficient alternatives have been proposed, the continual reassessment method (2) (CRM) being one example, but most of the investigators prefer to use the 3+3 design because it is simple to implement. A number of challenging problems in dose-finding have appeared recently where the use of the 3+3 design results in a very long or very inefficient trial or where it simply cannot be used. These problems include trials with delayed outcomes, studies of drug combinations, dose-finding based on toxicity scores, and dose-finding methods for cytostatic agents. We review some of these methods below.

Model-Based Designs for Phase I

Several model-based designs have appeared in the literature. Unlike the 3+3 design, these model-based designs base dose-escalation strategies on the accruing data through parameter estimation. One of the most used among such design is the CRM (2). The overall goal of a CRM procedure for dose escalation is to find the MTD as efficiently and safely as possible. The original CRM proposal used a simple parametric model to characterize the relationship between dose and the risk of a pre-defined dose-limiting toxicity (DLT). One then uses prior distributions on model parameters and results from patients treated in the current study to update the parameters and, thereby, the estimated relationship of dose to DLT risk.

In response to some criticisms, modified versions of the CRM appeared. Current versions run more or less as follows. The study investigators pick a target level of risk of DLT. This target, often around 30%, will reflect the investigators' subjective assessment of the therapeutic window of the treatment. In other words, the goal is to find thedose or dose level that is associated with the target risk of DLTs. This dose or dose level is sometimes called the MTD or the recommended phase II dose (RP2D). Traditionally, anticancer drugs have had a narrow dose range within which one might feel that the chance of benefit might offset the risk of serious or severe toxicity. Under these assumptions, a target risk of 20% to 33% may make sense.

After choosing a target risk, the investigators need to consider and set the dose levels to evaluate in the study. We use the term dose level as a generic term that includes doses of single agents or different dose combinations. Generally, the phase I protocols include specific discrete dose levels, although one may instead wish to consider a range and allow evaluation of any dose within this range. The latter approach will generally require a parametric model. An additional design requirement is the number of patients to treat at a given dose or dose level before deciding to escalate the dose for the next cohort. Typically, cohorts will consist of one to three patients.

Finally, the protocol needs a rule to indicate when to stop the study. One possibility is to stop when one has treated a pre-specified number of patients at a given dose level and the algorithm would assign that same dose to the next patient. For example, one may want to stop the study if one has treated six patients at dose level x and, based on all available data, the risk of DLT at dose level x is closer to the target than it is for any other dose level under evaluation, so the next patient would receive dose level x. This dose level would be the MTD or RP2D.

Babb, Rogatko, and Zacks (3) proposed a dose-escalation algorithm that, like the CRM, seeks to reach the MTD quickly but penalizes doses with predicted DLT risks above some threshold. This approach differs from the CRM as the CRM uses a penalty for each dose level based on the distance of its associated DLT risk from the target, regardless of whether it is above or below the target. Babb et al. called their algorithm EWOC for escalation with overdose control.

Newer agents, rather than attacking any cell that is dividing, say, have intracellular targets that appear associated with specific cancers. These targets may be altered signaling pathways or the state of methylation of genes in the cancer cells. While not free of untoward side effects, these newer agents have a very different toxicity profile than the earlier cytotoxic agents (4). A dose-finding trial for these agents can be based on a biomarker of an efficacy outcome that is often continuous (5), or both efficacy and toxicity can be used to determine the best dose (6). If outcome is binary dose-escalation schemes like the CRM are still useful.

Trials with Delayed Outcome

Dose limiting toxicity in phase I trials in solid tumors is often defined based on one cycle of therapy that usually lasts 3-4 weeks. In hematology oncology trials, follow-up for toxicity is usually 4-6 weeks. Another setting where long term toxicities are likely is radiation therapy trials. A number of methods have appeared for such trials that use information from all patients, not only patients who have completed follow-up. Cheung and Chappell (7) proposed a time-to-event modification of the CRM (2), called TITE-CRM. Ivanova et al. (8) proposed a similar frequentist method. The method of Bekele et al. (9) also uses partial information from patients to estimate the dose-toxicity relationship. Additionally, it prescribes if the next patient can be enrolled right away or if more time is needed to collect follow-up data on already enrolled patients. The method of Ivanova et al. (8) and its variations have been used in a number of trials at the Lineberger Comprehensive Cancer Center (LCCC) (10).

Dose-finding with Multiple Agents

It is common in oncology to treat patients with drug combinations. A method in Thall et al. (11) was one of the first dose-finding methods for drug combinations and has the ambitious goal of finding the whole contour of maximum tolerated drug combinations. The method is based on parametric functions that characterize each single agent's dose-DLT relationship and a model for the interaction. Other methods have a goal of finding several maximum tolerated drug-combinations, one for each dose of one of the agents (12,13), or a single maximum tolerated combination by searching over a grid of possible combinations (14,15,16). Though a number of methods have been developed for finding at least one maximum tolerated combination, nevertheless the method of choice to date has been to specify the sequence of drug combinations such that the MTD rate is monotonically increasing and use a single dimensional method, such as the 3+3 design or the CRM (2), to find the MTD.

Dose-finding Based on Toxicity Score

Toxicity in oncology is measured on a scale from 0 to 5. The DLT is typically defined as study drug related non-hematological toxicity of grade 3 or higher or study drug related hematological toxicity of grade 4 or higher. Bekele and Thall (17) used a sophisticated weighting system and ordinal toxicity variables with up to four categories. They proposed a Bayesian design for such trials. Some authors extended this idea and proposed to define a measure of toxicity called the toxicity index. The MTD is then defined as the dose with the toxicity index equal to a certain value (18,19). We are not aware of trials that used complex toxicity scoring. Our examples, two trials conducted at the LCCC (20,21), used a simpler approach where the MTD was defined based on DLT rate as well as on the rate of dose reductions. At the request of the investigators a rather simple method similar to the 3+3 design was used for assignment (22). Dose reduction rate was included in the definition of the MTD because the investigators did not want to recommend a dose that is likely to be reduced later in a cycle. The outcome of each patient was classified as DLT or toxicity that was not a DLT but caused a dose reduction or no toxicity. Since DLTs are more detrimental, they were scored twice as much compared to toxicities that cause dose reductions. The MTD was defined as the dose where

$$\Pr\{\text{DLT}\} + 0.5 * \Pr\{\text{dose reduction}\} = 0.25.$$

Safety Monitoring for Expansion Cohorts

Commonly, phase I studies in cancer include expansion cohorts. These are groups of patients who receive the putative recommended phase II dose or MTD after the dose escalation part of the study has determined the recommended dose. The goal of treating this group of patients may be to evaluate the drug in more patients and estimate the risk more precisely, to examine differences between men and women, or to get preliminary efficacy information. As long as the patients enrolled in the expansion cohort are roughly similar to the patients who participated in the dose-escalation part of the study (i.e., the expansion

patients satisfy the same eligibility criteria), it is reasonable to consider the information from the prior patients to apply to the expansion cohort.

Of particular relevance for the expansion cohort is the safety information from the dose-escalation part of the study. Since patients in this cohort will receive the MTD determined earlier in the study, one has preliminary information about the risk of a DLT. One can use this preliminary information as a basis for safety monitoring rules for the expansion cohort. For example, we might consider that six patients received the MTD prior to starting the expansion phase. Assuming a uniform prior (i.e., beta(1,1)) for the risk of DLT, one DLT out of six patients leads to a beta(2, 6) posterior probability distribution for the risk of DLT. If the target risk of DLT during dose escalation was 30%, then a reasonable safety monitoring rule would stop treating patients if it appears that the risk is higher than that target. Following this logic, one might want to stop treating patients if the posterior probability is 75% or more (i.e., 3:1 odds or more) that the risk of DLT is 0.3 or higher. This probability is easy to compute, since the posterior probability follows a beta distribution, and most statistical software packages have functions for evaluating probabilities for a beta distribution.

Phase II Methods in Oncology Trials

Review of Methods for Phase II Trials in Oncology

The general purpose of phase II studies in oncology is to get an early indication of activity of the treatment. In the past, the primary endpoint in oncology was related to tumor shrinkage. That is, if the treatment would lead to reduction in the size of the tumor within a short period of treatment, then the treatment might provide long-term benefit. Ultimately, the goal is to extend life, and tumor shrinkage is an early sign that the treatment might help eradicate disease. Unfortunately, tumor shrinkage does not always correlate with extended survival, and newer targeted agents may slow tumor growth—at least initially—rather than shrink tumors within a couple of months. As a result of these considerations, especially in the era of evaluating agents that target specific molecular pathways thought to be involved in cancer growth, many phase II studies use other endpoints than tumor shrinkage. For example, delaying disease progression may indicate drug activity, so the primary endpoint may be the probability of being alive and free of progression (i.e., progression-free survival) at some specific time after starting treatment, such as six months. Mick et al. (23) consider the ratio of the current time to progression to the patient's previous time to progression as a measure of benefit. Rosner et al. (24) propose the use of randomized discontinuation designs when tumor growth rates exhibit heterogeneity and the proposed treatment may slow or stop tumor growth but not shrink the tumors immediately.

Many patients, including some with metastatic disease, experience slow growth of their tumors. Heterogeneity in tumor growth across patients may make it difficult to tell if longer than expected times-to-progression are a result of the treatment or disease heterogeneity. Thus, phase II studies with progression-free survival as the primary endpoint should be randomized. A subcommittee of the U.S. National Cancer Institute has proposed guidelines for phase II study designs in oncology, in an effort to improve the state of the science (25).

These guidelines focus on phase II endpoints and reflect the committee's opinions regarding endpoint-appropriate designs.

Most of the oncology phase II trials are single arm and use either Simon's two stage design (26) or Gehan's design (27). Evidence shows (28) that, in recent years, Simon's design is the most frequently used design for phase II trials. Simon's design is a two-stage design with the possibility to stop for futility after stage one. Usually the design that yields the minimum total sample size, the minimax design, or the design that yields the smallest sample size under the null hypothesis, the optimal design, is used. Jung et al. (29) recommended considering not only the minimax and the optimal design, but all designs that minimize the weighted average of the total sample size and the expected sample size under the null hypothesis. This method yields a number of useful designs for an investigator to choose from, in particular, they offer a wider selection of stage one sample sizes. Other extensions of Simon's design include the three stage design (30), the Fleming's two-stage design (31) that allows stopping for either efficacy and futility, the two-stage design with ordinal outcomes (32, 33), and a two-stage design with two arms (34). Table 1 displays various extensions of the Simon's two-stage design that have been proposed for phase II oncology trials.

Two-stage Design with Ordinal Outcomes

As the goal of a phase II trial is to quickly screen new agents, the concept of looking at several outcomes at the same time is very appealing. Some investigators proposed looking at complete and partial responses (32), some proposed looking at tumor response and disease control (33) (defined as tumor response or stable disease), some at tumor response (or disease control) and progression free survival (35, 36). The treatment is considered promising if it improves at least one of the outcomes. To give more details, let p_T and p_D, $p_T \leq p_D$, denote the probability of tumor response and disease control in the population, respectively, and p_{0T} and p_{0D}, $p_{0T} \leq p_{0D}$, denote the null probabilities of tumor response and disease control. Define the two null hypothesis H_{0T}: $p_T \leq p_{0T}$ versus H_{1T}: $p_T > p_{0T}$, and H_{0D}: $p_D \leq p_{0D}$ versus H_{1D}: $p_D > p_{0D}$. Consider testing the intersection of these two hypothesis:

$$H_0 : H_{0T} \cap H_{0D} \text{ versus } H_1 : H_{1T} \cup H_{1D}.$$

The treatment is considered promising if either tumor response rate or disease control rate is promising. The treatment is not considered promising if both tumor response and disease control rates are low. Lu et al. (32) and Ivanova et al. (33) describe how to construct Simon-like two-stage designs for testing two components of the ordinal outcome, e.g. tumor response and disease control. Zee et al. (35) and Sun et al. (36) describe how to test a binary endpoint and a survival endpoint simultaneously in a two-stage trial.

Stopping Rules for Phase II Trials

While the purpose of phase II studies is to determine if the treatment has some activity against the disease, there may not be a lot of experience treating patients with the therapy. Thus, monitoring the phase II patients for safety is often appropriate. Monitoring toxicity on

a continuous basis, that is, being able to stop the study at any point, provides the best protection against observing an excessive number of toxicities and therefore is preferable to multi-stage stopping rules. Ivanova, Qaqish, Schell (37) illustrated that the Pocock-type (38) stopping boundary allows stopping the trial as early as possible, if the toxicity rate is high, and therefore prevents treating too many patients on a regimen that is not safe.

Geller et al. (39) proposed a Bayesian continuous stopping rule for phase II trials. Bayesian stopping rules for toxicity monitoring in phase II are similar to the Bayesian monitoring rules discussed earlier in the section relating to phase I studies. These rules would call for stopping the treatment if more patients are experiencing serious adverse events than one might have anticipated before starting the study. Of course, the procedure we discussed in the phase I section assumed that one has reduced the safety outcome to a simple yes or no event and that one has a target risk one does not want to exceed. For example, one might assume, based on earlier phase I studies, that the risk of serious adverse events is around 30%. Since phase I studies typically treat relatively few patients, there is likely uncertainty about the true risk. One might characterize this uncertainty with a prior distribution for the risk, such as a beta(6, 14) distribution. This prior corresponds to a mean risk of serious adverse events of 30%, with 90% probability that the risk is between 15% and 48%. If one has greater certainty about the risk, one should use a different prior distribution. The next step is to decide on the level of certainty that one requires to stop treating future patients. For example, one might want to be at least 80% certain that the risk of a serious adverse event is greater than 30%, corresponding to 4:1 odds. Often, the protocol will contain tables that illustrate the stopping rules and the average behavior of the rules under different scenarios corresponding to different underlying “true” risks. The final rules should reflect the concerns and judgment of the clinical investigators and the statisticians.

Phase II Trials with Two Arms

Kepner (34) described how to compute two-stage designs with the possibility of stopping for efficacy and/or futility after stage one when the outcome is binary. As it is time consuming to search for all possible designs, Kepner (34) proposed to search for designs with the first stage sample size that is the closest to one half of the total sample size.

Bayesian Phase II Designs

Phase II designs have typically been based on a frequentist, hypothesis-testing framework. For example, the Simon two-stage design seeks to minimize the sample size under the null hypothesis while achieving pre-specified size under the null and power to detect some alternative hypothesis. Several investigators have proposed phase II designs that rely on Bayesian computation. Cook and Johnson (40) proposed designs for comparing two hypotheses but from a Bayesian perspective. They use the Bayes factor, a measure of the strength of evidence in the data in favor of one hypothesis over another, for comparing the two hypotheses of interest. They develop this procedure using prior distributions based on alternatives that yield efficient designs.

Other designs in the literature consider stopping the study when there is high posterior probability that a key parameter, such as the probability of tumor shrinkage, exceeds a pre-

specified threshold. For example, Thall and Simon (41) propose a single-arm design that sets up decision rules based on posterior probabilities relating a new treatment to the assumed efficacy of a current standard treatment. Aside from attaching uncertainty to the measure of efficacy (e.g., response probability) of the experimental treatment, they also allow for uncertainty about the efficacy of the standard. Even though patients in the study only receive the new treatment and no new information regarding the current standard will arise from the study, accounting for uncertainty about the standard's efficacy via a prior distribution adds more variation and improves the frequentist characteristics of the design. Tan and Machin (42) describe a two-stage design that uses posterior probabilities to decide whether to proceed from the first stage to the second. Sambucini (43) modifies the Tan and Machin design to use predictive probabilities rather than posterior probabilities. By using predictive probabilities, the design accounts for variation in future observations. Furthermore, Sambucini considers a range of alternatives, via a design prior, rather than fixing the “true” underlying response probability at a single point.

Lee and Liu (44) proposed a related procedure that also bases decisions on predictive probabilities, rather than posterior probabilities. Their design will ensure control of Type I and Type II error probabilities. Furthermore, unlike the Tan and Machin approach, Lee and Liu's proposal allows one to monitor the study as often as one needs. Continual monitoring has the potential to lead to smaller expected sample sizes under the null hypothesis, compared to a design with just one or two interim futility analyses.

Confirmatory Oncology Trials

Traditionally, phase III in oncology drug development has been the first randomized controlled study phase. Some of the reasons for this phenomenon include the dose-dependent nature of toxicity for most oncology agents, the acceptance of toxic treatments by oncologists and by patients, and belief that small trials with an observed response rate as a primary endpoint have sufficient information to guide further development of a clinical program. Insufficient exploration of safety and efficacy of cancer drugs in early phases of development together with the complex life-threatening nature of the disease has led oncology drug development to have a high rate of failure in confirmatory trials (45). In order to compensate for some oncology development shortcuts, to improve the chance of a success, and to expedite time-to-market, some phase III trials have been employing designs with adaptive features such as the possibility to stop early for futility or efficacy, sample size re-estimation, implementation of enrichment strategies, etc. Stopping for efficacy during an interim analysis is not commonly considered because the full sample size is usually required in order to satisfy regulatory requirements for the adequacy of the safety database in the overall development program. But if the treatment effect is very strong, ethical considerations may warrant early termination of the study (46). Bretz et al. give a nice overview of adaptive designs used in confirmatory clinical trials (47).

Seamless phase II/III designs have become more popular in oncology drug development (48,49,50,51). Such designs aim to reduce the overall sample size by allowing the data from phase II patients to be used in phase III analysis (inferentially seamless) (48,49) or/and eliminating the time between phases, which results in a shorter total drug development time

(operationally seamless). Just as there are a number of phase II designs, there are a number of corresponding phase II/III designs. More on different seamless designs in oncology and design efficiency in phase II/III trials can be found in Korn et al (52).

Available Software

MD Anderson Comprehensive Cancer Center provides a site with software to implement a number of designs for oncology trials: https://biostatistics.mdanderson.org/SoftwareDownload/. All admissible Simon's (26) and Fleming's (31) designs can be generated using the software available at the Lineberger Comprehensive Cancer Center site at http://cancer.unc.edu/biostatistics/program/ivanova/. This software can also generate the two-stage designs with ordinal outcome (33) and the Pocock boundary for continuous toxicity monitoring in a single arm phase II study (37). Two-stage, two-arm study designs for binary outcomes can be generated using software at www.cryptnet.net/kepner.

Discussion

Oncology drug development has been a tremendous challenge. Cancer clinical studies are typically lengthy and costly; yet, the failure rate is high. For example, only 34% of phase III oncology trials with results reported from 2003 through 2010 were successful (45). While insufficient understanding of the disease (such as pathway targets) and effective candidate therapeutics certainly contribute to the high failure rate, it has been argued by many that more robust clinical development strategies need to be in place to increase the probability of success in oncology drug development.

In recent years, a considerable amount of scientific work has focused on early phase oncology development. While the “3+3” design remains the most used phase I design, according to a literature search by Rogatko et al. (53) and Le Tourneau et al. (54), the CRM design and its modifications are making their way to trials run by pharmaceutical industry (55,56). The 3+3 design is inefficient in establishing the dose that meets a specific target toxicity level and may involve an excessive number of escalation steps, leading to a large proportion of patients treated at sub-therapeutic doses. Additionally, the 3+3 design has a low probability of selecting the true MTD (57) and yields high variability in MTD estimates (58). Model-based methods that use all toxicity information accumulated during the trial achieve better estimates of the target probability of dose-limiting toxicity at the recommended dose without treating too many patients at sub-therapeutic doses, but implementation of these designs is not as simple as for the 3+3 design. There is a need for upfront planning, expedited collection of data from each patient or cohort to fit the model, biostatistics expertise, and software to perform model fitting in real time. Following the considerable amount of criticism received by the original version of the CRM, the algorithm has since been substantially modified and refined, both in its theoretical development and in its practical implementation. Seamless designs including phase I/II (mostly operationally seamless) have become more popular in the pharmaceutical industry. The first part (phase I) uses a model-based design to find the MTD, and the second part (phase II) treats a cohort of additional patients at the MTD to evaluate efficacy and further study safety in the same population or in a particular indication proposed for further development. The setting is

similar to safety monitoring for expansion cohorts. Drug combination studies are common in oncology and require special consideration for identifying an optimal dose combination. Phase I trials with drugs expected to produce delayed or cumulative toxicities should incorporate specific dose-finding methods and account for delayed or cumulative toxicities.

A paradigm shift in the past decade from traditional cytotoxic agents to molecular targeted therapies (from drugs primarily targeting DNA to those primarily targeting cellular signaling) has occurred. Given the different mechanisms and toxicities of these agents, drug development methodology may need to change. The Task Force on Methodology for the Development of Innovative Cancer Therapies (MDICT) was appointed by the NDDO Research Foundation (www.nddo.org) for the purpose of addressing methodological issues created by this shift. The MDICT published its recommendations on Phase I and II studies of targeted anticancer therapy but has not provided guidance on statistical methods specific for molecularly targeted compounds, acknowledging that “toxicity remains the most commonly used information upon which decisions are made for the recommended Phase II dose of targeted agents” (59,60). The phase II study design is a critical component of the clinical development strategy and has been one of the focal points in recent literature of oncology clinical development strategy. In this article, we reviewed some of the recent developments in Phase II study designs in oncology. Traditionally, single-arm study designs have been the mainstream in phase II oncology development, but because of patient selection and other biases associated with single-arm studies there has been a gradual shift to randomized designs for phase II. Yet, for certain mono-therapy trials with tumor response rate as the primary endpoint, single-arm designs are still acceptable (25). For single-arm studies, Simon's two-stage design and its various extensions have been commonly used due to its simplicity and its built-in mechanism to stop early due to futility. Assuming the same Type I and II errors, the design is optimal in the sense of minimizing the expected sample size under the null hypothesis. In practice, however, to achieve the optimal expected sample size, the sponsor will need to put the enrollment on hold after the first stage patients are enrolled until their data are available for analysis. One approach to avoid halting enrollment with a two-stage design was given by Herndon (61).

Because of the inherent drawbacks of single-arm designs, randomized designs are now being increasingly used in phase II oncology development, particularly when a time-to-event outcome, such as progression-free survival, is the primary endpoint. Randomized designs offer more objective and robust comparisons between experimental and control regimens, thus enabling better commercial decision making (so called Go/No-Go decision making). The price, however, is that the size of phase II may be much larger than single-arm studies to achieve reasonable study power, which could be a challenge for the “speed to market” strategy pursued by many sponsors in the increasingly competitive oncology market. On the other hand, there are rich historical data on the control regimen in many situations that could and should be leveraged when designing randomized phase II studies. Bayesian methods that incorporate historical control data into randomized studies have been proposed in phase II oncology development (62). In phase II development, there remains a huge unmet need for better, more efficient designs that could enable better and faster Go/No-Go decision making while controlling the cost. Some have advanced the idea of incorporating decision theory into the design of phase II studies of cancer therapies (63,64). Potential areas for

research include better, faster, and cheaper surrogate endpoints for survival, treatment-selection biomarkers that predicts a large treatment effect, etc.

While adaptive designs have become more popular in every stage of drug development, phase III designs are more conservative, given the regulatory requirements for confirmatory trials. Seamless phase II/III trials are getting some recognition and gaining popularity in oncology drug development. Because the regulatory issues are similar for seamless phase II/III trials as for phase III adaptive trials, the FDA guidance does not distinguish them. While regulators suggest limiting the use of adaptive designs at the confirmatory stage, they encourage further exploration of new methods, including adaptive designs in early phase of drug and biologic development (46).

Heterogeneity of disease is a widely acknowledged problem in cancer, and very difficult to address in the course of a clinical trial. It is the “next big problem” that oncology trial design has to tackle; perhaps designs used in BATTLE (65), I-SPY 1 (66), and I-SPY 2 (67) trials (www.ispy2.org) that pair oncology therapies and biomarkers point the way. Many newer designs, while superior in some ways to currently used designs, do not find their way into common usage right away. There are various reasons, including the need for specialized software, difficulty for non-statisticians to understand the methods, and reluctance on the part of government agencies to accept novel designs. In this paper, we have presented many new innovations in the design of clinical studies to hasten their acceptance and adoption.

Acknowledgments

Rosner's work was supported by the cancer center's core grant: P30 CA006973.

References

1. Korn EL, Midthune D, Chen TT, Rubinstein LV, Christian MC, Simon RM. A comparison of two phase I trial designs. Statistics in Medicine. 1994; 13:1799–1806. [PubMed: 7997713]
2. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: A practical design for phase I clinical trials in cancer. Biometrics. 1990; 46:33–48. [PubMed: 2350571]
3. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. Stat Med. 1998; 17:1103–1120. [PubMed: 9618772]
4. Korn EL. Nontoxicity endpoints in phase I trial designs for targeted, non-cytotoxic agents. Journal of the National Cancer Institution. 2004; 96:977–978.
5. Hoering A, Lebilanc M, Crowley J. Seamless phase I-II trial design for assessing toxicity and efficacy for targeted agents. Clinical Cancer Research. 2011; 17:640–646. [PubMed: 21135145]
6. Ivanova A, Kim S. Dose-finding for binary ordinal and continuous outcomes with monotone objective function: a unified approach. Biometrics. 2009; 65:307–315. [PubMed: 18479486]
7. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. Biometrics. 2000; 56:1177–1182. [PubMed: 11129476]
8. Ivanova A, Flournoy N, Chung Y. Cumulative cohort design for dose-finding. Journal of Statistical Planning and Inference. 2007; 137:2316–2317.
9. Bekele BN, Ji Y, Shen Y, Thall PF. Monitoring late-onset toxicities in phase I trials using predited risks. Biostatistics. 2008; 9:442–457. [PubMed: 18084008]
10. Foster MC, Amin C, Voorhees PM, et al. A phase I, Dose Escalation Study of Clofarabine in Combination with Fractionated Gemtuzumab Ozogamicin in Patients with Refractory or Relapsed Acute Myeloid Leukemia. Leukemia and Lymphoma. 2012 in press.

11. Thall P, Millikan R, Mueller P, Lee S. Dose finding with two agents in phase I oncology trials. Biometrics. 2003; 59:487–496. [PubMed: 14601749]
12. Ivanova A, Wang K. A nonparametric approach to the design and analysis of two-dimensional dose-finding trials. Statistics in Medicine. 2004; 23:1861–1870. [PubMed: 15195320]
13. Wang K, Ivanova A. Two-dimensional dose finding in discrete dose space. Biometrics. 2005; 61:217–222. [PubMed: 15737096]
14. Conaway MR, Dunbar S, Peddada SD. Designs for single- or multiple-agent phase I trials. Biometrics. 2004; 60:661–669. [PubMed: 15339288]
15. Yuan Y, Yin G. Sequential continual reassessment method for two-dimensional dose finding. Statistics in Medicine. 2008; 27:5664–5678. [PubMed: 18618901]
16. Braun TM, Wang S. A hierarchial bayesian design for phase I trials of novel combinations of cancer therapeutic agents. Biometrics. 2010; 66:805–812. [PubMed: 19995354]
17. Bekele BN, Thall FT. Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. Journal of American Statistical Association. 2004; 99:26–35.
18. Yuan Z, Chappell R, Bailey H. The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach. Biometrics. 2007; 63:173–179. [PubMed: 17447942]
19. Chen Z, Krailo MD, Azen SP, Tighiouart M. A novel toxicity scoring system treating toxicity response as a quasi-continuous variable in phase I clinical trials. Contemporary Clinical Trials. 2010; 31:473–482. [PubMed: 20609419]
20. Raftery L, Tepper JE, Goldberg RM, et al. A Two-cohort phase 1 study of weekly oxaliplatin and gemcitabine, then oxaliplatin, gemcitabine, and erlotinib during radiotherapy for unresectable pancreatic carcinoma. American Journal of Clinical Oncology. 2012 in press.
21. O'Neil BH, Raftery L, Calvo BF, et al. A phase I Study of Bortezomib in Combination with Standard 5-fluorouracil and External Beam Radiotherapy for the Treatment of Locally Advanced or Metastatic Rectal Cancer. Clinical Colorectal Cancer. 2010; 9(2):119–125. [PubMed: 20378507]
22. Ivanova A. Escalation, up-and-down and A+B designs for dose-finding trials. Statistics in Medicine. 2006; 25:3668–3678. [PubMed: 16381057]
23. Mick R, Crowley JJ, Carroll RJ. Phase II Clinical Trial Design for Noncytotoxic Anticancer Agents for Which Time to Disease Progression Is the Primary Endpoint. Controlled Clinical Trials. 2000; 21:343–359. [PubMed: 10913809]
24. Rosner GL, Stadler W, Ratain MJ. Randomized discontinuation design: application to cytostatic antineoplastic agents. J Clin Oncol. 2002; 20:4478–4484. [PubMed: 12431972]
25. Seymour L, Ivy SP, Sargent D, et al. The Design of phase II Clinical Trials Testing Cancer Therapeutics: Consensus Recommendations from the Clinical Trial Design Task Force of the National Cancer Institute Investigational Drug Steering Committee. Clinical Cancer Research. 2010; 16:1764–1769. [PubMed: 20215557]
26. Simon R. Optimal two-stage designs for phase II clinical trials. Controlled Clinical Trials. 1989; 10:1–10. [PubMed: 2702835]
27. Gehan E. 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. 1961; 13:346–353. [PubMed: 13704181]
28. Thezenas S, Duffour J, Culine S, Kramar A. Five-year change in statistical designs of phase II trials published in leading cancer journals. European Journal of Cancer. 2004; 40:1244–9. [PubMed: 15110889]
29. Jung SH, Lee T, Kim KM, George SL. Admissible two-stage designs for phase II cancer clinical trials. Statistics in Medicine. 2004; 23:561–569. [PubMed: 14755389]
30. Cheng T. Optimal three-stage designs for phase II cancer clinical trials. Statistics in Medicine. 2007; 16:2701–2711.
31. Fleming TR. One-Sample multiple testing procedure for phase II clinical trials. Biometrics. 1982; 38:143–151. [PubMed: 7082756]
32. Lu Y, Jin H, Lamborn KR. A design of phase II cancer trials using total and complete response endpoints. Statistics in Medicine. 2005; 24:3155–3170. [PubMed: 16189806]

33. Ivanova A, Monaco J, Stinchcombe T. Efficient designs for phase II oncology trials with ordinal outcome. Statistics and Its Interface. 2012; 5:463–469.
34. Kepner JL. On group sequential designs comparing two binomial proportions. Journal of Biopharmaceutical Statistics. 2010; 20(1):145–59. [PubMed: 20077254]
35. Zee B, Melnychuk D, Dancey J, Eisenhauer E. Multinomial phase II cancer trials incorporating response and early progression. Journal of Biopharmaceutical Statistics. 1999; 9:351–363. [PubMed: 10379698]
36. Sun LZ, Chen C, Patel K. Optimal two-stage randomized multinomial designs for phase II oncology trials. Journal of Biopharmaceutical Statistics. 2009; 19:485–493. [PubMed: 19384690]
37. Ivanova A, Qaqish BF, Schell MJ. Continuous toxicity monitoring in phase I trials in oncology. Biometrics. 2005; 61:540–545. [PubMed: 16011702]
38. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977; 64:191–199.
39. Geller, NL.; Follmann, DF.; Leifer, ES.; Carter, SL. Design of early trials in peripheral blood stem cell transplantation: a hybrid frequentist-Bayesian approach. In: Geller, NL., editor. Advances in Clinical Trial Biostatistics. New York and Basel: Marcel Dekker, Inc; 2005. p. 40-52.
40. Johnson VE, Cook JD. Bayesian design of single-arm phase II clinical trials with continuous monitoring. Clin Trials. 2009; 6:217–226. [PubMed: 19528131]
41. Thall PF, Simon R. Practical Bayesian guidelines for phase IIB clinical trials. Biometrics. 1994; 50:337–349. [PubMed: 7980801]
42. Tan SB, Machin D. Bayesian two-stage designs for phase II clinical trials. Stat Med. 2002; 21:1991–2012. [PubMed: 12111883]
43. Sambucini V. A Bayesian predictive two-stage design for phase II clinical trials. Stat Med. 2008; 27:1199–1224. [PubMed: 17763528]
44. Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clin Trials. 2008; 5:93–106. [PubMed: 18375647]
45. Sutter, S.; Lamotta, L. Cancer drugs have worst phase III track record. Internal Medicine News. 2011. available at http://www.internalmedicinenews.com/news/oncology-hematology/single-article/cancer-drugs-have-worst-phase-iii-track-record/32e1dca10b.html
46. FDA. Guidance for Industry. Adaptive Design Clinical Trials for Drugs and Biologics. 2010. http://www.fda.gov/downloads/DrugsGuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf
47. Bretz F, Koenig F, Brannath W, Glimm E, Posch M. Adaptive designs for confirmatory clinical trials. Statistics in Medicine. 2009; 28:1181–1217. [PubMed: 19206095]
48. Bretz F, Schmidli H, Konig F, et al. Confirmatory seamless phase II/III clinical trials with hypothesis selection at interim: General concepts. Biom J. 2006; 48:623–634. [PubMed: 16972714]
49. Stallard N. A confirmatory seamless phase II/III clinical trial design incorporating short-term endpoint information. Statistics in Medicine. 2010; 29:959–971. [PubMed: 20191605]
50. Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. N Engl J Med. 2011; 364:1817–1825. [PubMed: 21561347]
51. Bookman MA, Brady MF, McGuire WP, et al. Evaluation of new platinum-based treatment regimens in advanced-stage ovarian cancer: A phase III trial of the Gynecologic Cancer InterGroup. J Clin Oncol. 2009; 27:1419–1425. [PubMed: 19224846]
52. Korn E, Friedlin B, Abrams J, Halabi S. Design issues in randomized phase II/III trials. Journal of Clinical Oncology. 2012; 30(6):667–671. [PubMed: 22271475]
53. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A. Translation of innovative designs into phase I trials. J Clin Oncol. 2007; 25(31):4982–4986. [PubMed: 17971597]
54. Le Tourneau C, Lee JJ, Siu L. Dose escalation methods in phase I cancer clinical trials. J Natl Cancer Inst. 2009; 101(10):708–720. [PubMed: 19436029]
55. Neuenschwander B, Branson M, Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. Statistics in Medicine. 2008; 27:2420–2439. [PubMed: 18344187]

56. Bailey S, Neuenschwander B, Laird G, Branson M. A Bayesian case study in oncology phase I combination dose-finding using logistic regression with covariates. Journal of Biopharmaceutical Statistics. 2009; 19:469–484. [PubMed: 19384689]
57. Thall PF, Lee SJ. Practical model-based dose-finding in phase I clinical trials: Methods based on toxicity. International Journal of Gynecological Cancer. 2003; 13(3):251–261. [PubMed: 12801254]
58. Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. Statistics in Medicine. 1995; 14:1149–1161. [PubMed: 7667557]
59. Booth CM, Calvert AH, Giaccone G, Lobbezoo MW, Seymour EA, Eisenhauer EA. Endpoints and other considerations in phase I studies of targeted anticancer therapy: Recommendations from the Task Force on Methodology for the Development of Innovative Cancer Therapies (MDICT). Eur J Cancer. 2007; 44(1):19–24. [PubMed: 17890079]
60. Booth CM, Calvert AH, Giaccone G, Lobbezoo MW, Eisenhauer LK, Seymour LK. Design and conduct of phase II studies of targeted anticancer therapy: Recommendations from the Task Force on Methodology for the Development of Innovative Cancer Therapies (MDICT). Eur J Cancer. 2007; 44(1):25–9. [PubMed: 17845846]
61. Herndon JE. A design alternative for two-stage, phase II, multicenter cancer clinical trials. Controlled Clinical Trials. 1998; 19:440–450. [PubMed: 9741865]
62. Mugglin A. Case Study: Using Historical Information to Augment Prospective Control. FDA/ Industry Workshop. 2008
63. Stallard N, Thall PF, Whitehead J. Decision theoretic designs for phase II clinical trials with multiple outcomes. Biometrics. 1999; 55:971–977. [PubMed: 11315037]
64. Ding M, Rosner GL, Müller P. Bayesian optimal design for phase II screening trials. Biometrics. 2008; 64:886–894. [PubMed: 18162113]
65. Zhou X, Liu S, Kim ES, Herbst RS, Lee JJ. Bayesian adaptive design for targeted therapy development in lung cancer – a step toward personalized medicine. Clinical Trials. 2008; 5(3): 181–93. [PubMed: 18559407]
66. Esserman LJ, Berry DA, Cheang MC, et al. Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). Breast Cancer Res Treat. 2012; 132(3):1049–62. [PubMed: 22198468]
67. Barker AD, Sigman CC, Kelloff GJ, et al. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clin Pharmacol Ther. 2009; 86(1):97–100. [PubMed: 19440188]

Table 1

Extensions of the Simon's two-stage design for phase II oncology trials

Simon's two stage design	Extensions of the Simon's design
Two-stage	Three-stage (30)
Stopping for futility only	Stopping for futility and efficacy (31)
Binary endpoint	Ordinal outcome (32, 33)
Minimax or optimal	All admissible designs (29)
Single arm	Two-arm (34)