Implementing Adaptive Dose-Finding Designs in Oncology Clinical Trials

Bo Huang[1], Paul Bycott[1] and Enayet Talukder[1]
[1]Oncology Business Unit, Pfizer Inc.

Abstract

The main objective in phase I oncology trials is to identify a maximum tolerated dose for subsequent studies. Over the past two decades, extensive research has been conducted by statisticians and clinical researchers to create innovative dose finding designs that perform better than the standard 3+3 design, which often exhibits undesirable statistical and operational properties. However, clinical implementation and practical usage have been limited. This article provides some perspectives on implementing adaptive dose finding designs in oncology phase I trials. A case study is provided to illustrate why, how and when novel statistical methodologies are adopted. Operational considerations on how to effectively conduct such trials are discussed.

Key words: Dose Finding, Phase I, Adaptive Design, Oncology, MTD

1. Introduction

The primary objective of a phase I clinical trial in cancer or other life-threatening illnesses is to determine the maximum tolerated dose (MTD) or a recommended dose of a new treatment for subsequent clinical evaluation of efficacy. The 3+3 algorithm since its inception has been the standard design for oncology dose finding studies primarily because of its simplicity. Although a strict quantitative definition of the MTD is seldom acknowledged in clinical protocols particularly when the standard method is used (Storer, 1989), in most statistical literature it is defined as the dose for which the probability of the dose-limiting toxicity (DLT) is equal to a specified value θ_0, often (but not always) in the range between 0.20 and 0.33. In cancer phase I trials, participants are usually late-stage patients for whom most or all alternative therapies have failed. For such patients, toxicity may be severe before it is considered intolerable. Generally, grade III or higher toxicity of a cytotoxic agent is considered dose limiting, which means it is severe enough that the treatment must be at least temporarily discontinued.

The standard 3+3 design, despite its simplicity and popularity, often does not exhibit proper statistical properties. In particular, patients are more likely to be treated outside the potential therapeutic window and the final estimation of the MTD tends to be biased. The design is inefficient and fails to make use of all available data as escalation/de-escalation decisions are based only on the most current cohort of patients. In addition, the choice of the target rate of DLT is unclear and restricted. Lastly and importantly, the standard method cannot handle complicated situations (e.g. drug combination, multiple endpoints, toxicity grade information).

Correspondence to: Bo Huang, PhD., MS8260-2123, Eastern Point Road, Groton, CT 06340, U.S.A. Email: bo.huang@pfizer.com

Numerous alternative designs have been proposed for Phase I dose escalation trials that target to improve accuracy, efficiency and statistical validity, including the continual reassessment method (CRM) (O'Quigley et al., 1990), and its variants. However, based on a study on cancer phase I trial publications from 1991 to 2006 (Rogatko et al., 2007), an overwhelming 98.4% of the clinical trials (1,215 of 1,235 trials) followed variations of the standard method. Only 20 trials used Bayesian adaptive designs. These 20 clinical studies showed extensive lags between publication of the statistical paper and its translation into a clinical paper. Much work remains to be done to close the gap so that better methods can be used in clinical practice in order to improve cancer patients' survival and quality of life. It also entails thorough evaluation of the existing methods and to select the most suitable design for any specific trial.

We first give an overview of existing dose finding methods and designs in cancer research, and discuss the challenges in translating innovative designs into clinical trials in Section 2. One dose finding case study in cancer is presented in Sections 3. In Section 4, we share our experience in operation of conducting clinical trials when adaptive designs are used, followed by discussion in Section 5.

2. Dose-Finding Designs

Dose finding designs for cancer phase I trials generally fall into two classes, the algorithm-based designs and the model based designs. The algorithm-based designs (including the 3+3) assign patients to doses and select the MTD according to pre-specified algorithms based on the observed data. On the other hand, model-based designs assign patients to doses based on a dose-toxicity model to estimate the MTD associated with a target toxicity rate. In addition, some designs were created by incorporating features of algorithm-based decision rules into simple statistical models (or vice versa) so as to take advantage of benefits from both worlds. We'll give a brief overview of these designs.

2.1. Algorithm-Based Designs

Algorithm (rule)-based designs are typically called "up-and-down" designs that do not specify prior assumption on the dose-toxicity relationship. The main feature of this type of designs is that the decision rules are pre-specified before the actual observation of the toxicity data. Storer (1989) summarized a number of common up-and-down designs including the standard 3+3 designs and its variation. Lin and Shih (2001) and Ivanova (2006) proposed A+B designs that are more flexible and increase patient safety. Durham (1997) developed a "biased coin" design with decision not fixed and some type of stochastic rules added to dose escalation. Simon (1997)'s accelerated titration design is a popular rule-based design which allows intra-patient dose escalation and eventually recommends using a statistical model to select the MTD. The intra-patient dosing and the initial acceleration features are appealing as they could substantially reduce the study size and save resources. It is most useful for single-agent testing with short pharmacokinetic half-life and transient adverse events. In other cases (such as biologics) due to cumulative toxicities and patient drop-out, the potentially induced biases have limited its practical utility and may result in uninterpretable data. Another common algorithm-based design is the isotonic regression (Leung et al, 2001) to address the violation of monotonicity in toxicity due to small samples by using the pool-adjacent-violators-algorithm (PAVA).

The up-and-down designs are generally "short-memory" (Oron and Hoff, 2013) in that only the most recent cohort of patients is utilized for escalation and de-escalation decisions. Compared to some of the most advanced model-based methods, algorithm-based designs generally assign a higher percentage of patients to low doses that are potentially subtherapeutic and a lower percentage of patients at or near the true MTD or recommended dose for phase 2 trials. Furthermore, most algorithm-based designs were developed to handle simple dose-finding problems (i.e. single agent dose finding of the MTD using binary DLT data). Complicated situations such as heterogeneous population, drug combination, multiple-endpoint and toxicity grade information may require advanced model-based methods in order to be tackled adequately.

Ji et al. (2010) developed an algorithm-based design in which patients are assigned to dose levels according to the DLT outcomes at the current dose by calculating the toxicity probability intervals under the beta-binomial model. Liu and Yuan (2012) proposed an optimal interval design, where decision of dose assignment is determined by comparing the observed toxicity rate at the current dose with a prespecified optimal interval. Simulations have shown that the performance of these type of hybrid designs that incorporate features of both model-based and algorithm based designs are better than the 3+3 design and comparable to some model-based designs. Importantly, they maintain the strength of algorithm-based designs of operational simplicity and transparency in practice.

2.2. Model-Based Designs

Model-based dose finding designs are another class of phase I designs. The main characteristic of model-based designs is the application of a dose-toxicity model to estimate the MTD associated with a target toxicity rate θ_0. This type of design can be conveniently implemented in the Bayesian framework, with an initial estimation of the shape and scale of the dose-toxicity curve as the prior distribution, and a posterior distribution subsequently estimated after incorporating observed toxicity data at each dose into the model. The posterior is then evaluated to identify the dose closest the MTD as defined. The continual reassessment method (CRM) is a Bayesian parametric method proposed by O'Quigley et al. (1990) in which single-patient cohorts are successively assigned to the current posterior estimates of the MTD. It has gained widespread popularity by providing a more efficient and accurate MTD estimate. Due to criticism for its higher probability of treating patients above the MTD and longer duration, it was modified and refined in several subsequent papers (Faries, 1994; Goodman et al., 1995; O'Quigley and Shen, 1996) with safeguards to address practical, safety and ethical issues. The Escalation with Overdose Control (EWOC) design is a variation of the CRM method by adding an additional restriction to dose escalation to address the ethical concern of over-dosing (dose above the MTD). It requires the Bayesian posterior probability of over-dosing cannot exceed a certain threshold level in order to control the probability of testing and choosing a dose above the MTD.

Model-based dose finding designs are "long-memory" as they incorporate all toxicity data at each dose to the model, as opposed to the "short-memory" up-and-down designs that use only the most recent cohort or all patients at the current dose for escalation decision. The more efficient use of data via a statistical model for the dose-toxicity relationship could substantially enhance the operating characteristics of the design compared to the standard design. Theoretical work also demonstrates that the CRM and

EWOC designs possess sound statistical properties including consistency (the model selects the true MTD eventually) and coherence (escalation/de-escalation does not occur when current dose cohort is deemed toxic/safe) if the models meet certain regularity conditions (Shen and O'Quigley, 1996; Cheung, 2005; Huang and Chappell, 2008). In addition, dose escalation may be accelerated if one patient per dose level is implemented initially, but it may also deprive the study team of data on interpatient pharmacokinetic/pharmacodynamic variability. In clinical practice, it's common to require cohort sizes be 2-4 to address this limitation. Finally, simulations from Oron and Hoff (2013) suggest the class of CRM designs is sensitive to data in early cohorts due to the long-memory nature which may result in unexpected variability in trial performance. As the dosing scheme of model-based designs cannot be pre-specified in advance and involves prior and model assumptions on the unknown dose-toxicity curve, it is prudent to include some "gate-keeping" rules (e.g. no or limited skipping of dose) in the protocol to protect patients from unethical and incoherent decisions.

2.3. Designs for Complicated Dose-Finding Trials

As the landscape of cancer research evolves rapidly thanks to the emergence of molecular targeted therapies, some of the traditionally developed dose finding methods may no longer be adequate or efficient. New methods have been developed to meet the challenges.

Traditionally efficacy is not considered as a primary endpoint in a phase I trial and it is generally assumed that efficacy increases with dose until the MTD, same with toxicity. However, this assumption may not always hold. The therapies may be cytostatic in nature and efficacy may go up initially and level off afterwards or even go down. To select the optimal dose for subsequent trials, it is appealing to evaluate safety and efficacy simultaneously. Braun (2002) extended the CRM to the bivariate trial design and named it bCRM. Thall and Cook (2004) defined an acceptable dose combination based on trade-offs between the probabilities of treatment efficacy and toxicity. Although these designs are appealing and efficient, there has been limited use in practice as it's challenging to identify a reliable endpoint for clinical efficacy that can be evaluated in a timely manner.

With the recent wave of new molecular-targeted agents that are not cytotoxic in mechanism of action, the interest to identify a biological or pharmacological optimal dose is rising in recent years. Unlike clinical efficacy endpoints which usually require a longer time to evaluate, markers of target modulation/inhibition and PK data may be reported rapidly and the results could be made available prior to making the next dosing decision. In this case, the definition of RP2D should be based on toxicity, PK and PD biomarker data (Fox et al., 2002). If target or exposure saturation is achieved prior to reaching the MTD, escalation beyond that dose may induce excessive toxicity due to off-target effects at higher drug concentrations. To identify the optimal biologic dose (OBD) is an exciting but highly challenging goal, partly due to the lack of precise understanding of biology and clear and quantitative definition of target modulation in order to guide dose escalation. In addition, large interpatient variability is expected at small sample sizes at each dose level. Lastly availability of tumor tissues and reliable assays are required to measure the drug effect on the target. This may explain the stark contrast between academic enthusiasm and rare clinical adoption. Nonetheless, a number of designs were developed by incorporating a biologic or pharmacokinetic endpoint in order to improve the selection of the recommended dose for phase II (Piantadosi and Liu 1996; Mandrekar et al., 2007; Polly and Cheung, 2008).

Clinical trials for combination of agents have become increasingly common in recent years. A combination of drugs can target cancer cells that have differing drug susceptibilities, achieve a higher intensity of dose if the drugs have nonoverlapping toxicities and reduce the likelihood of drug resistance (Dancey, 2006). However, complexity of the design of a phase I trial increases exponentially with the number of different drugs included in the combination strategy. When drugs in combination have different mechanisms of action or nonoverlapping toxicities, the recommended dose for phase II for the combination may be near the recommended dose of each drug given as a single agent. However, as the biological effects of the combination may be quite complex and the PK/PD drug-drug interaction between the agents is largely unknown, it is often difficult to administer at the recommended dose of each drug given as a single agent. Unlike single-agent dose escalation where monotonicity is generally assumed to be true, in drug combination only partial ordering is known for the dose-toxicity relationship. For instance for dual-agent A (A_1, A_2…) +B (B_1, B_2…), ordering in toxicity holds for some dose combinations so that $A_1+B_1< A_2+B_2$ but unknown for others (e.g. A_2+B_1vs A_1+B_2). A number of escalation strategies have been proposed in the literature and were described nicely in Tourneau et al. (2009) and Harrington et al. (2013). For dual-agent combination, when escalation occurs on both agents (no agent is fixed), the rule-based 3+3 or A+B design can be extended to the 3+3+3 or A+B+C design. Some flexible but more complex model-based designs were also published with model parameters to account for the inherent complexity of drug combination in the dose-toxicity relationship (Kramar, et al., 1999; Thall et al., 2003; Huang et al., 2007; Yuan and Yin, 2008; Yin and Yuan, 2009). Choosing the suitable dose escalation strategy and the right doses remain a great challenge in the development of combination therapies in cancer, and it should be determined by the best possible scientific and clinical practice rationale.

2.4. Challenges in the Translation of Novel Designs to Clinical Trials

In the areas of Phase 2-4 studies, the scientific approaches guiding the statisticians and the clinicians are largely in harmony, where the use of randomization is the gold standard to objectively assess whether the experimental treatment works without introducing bias. This is not so, however, for Phase 1 dose-finding studies (O'Quigley, 2009). The statistical theory of experimental design, in the context of a dose-finding study, would in most cases aim at efficiency and precision on some hypothesized dose–toxicity curve. It is the reason that many statistical designs are proposed in a sense to increase efficiency and maximize the chance of patients being treated around the MTD (Huang and Chappell, 2008; Bailey, 2009). However, it is fair to say efficiency and precision is not yet a priority for clinicians in the design of their dose finding trials, and they are not yet fully on board with the statisticians in working on superior designs to the 3+3 method created nearly half a century ago. As stated in O'Quigley (2009), many clinicians, with the lack of probability background, will intuitively be in favor of the standard approach, for which (targeting the 33rd percentile) they expect at doses under the MTD to see zero toxicity for every three included patients, and at the MTD, exactly one out of three patients experiencing a toxicity. This is one of the reasons why the outdated and inadequate 3+3 design remains in common use. Therefore, in order to make progress on the application of better designs, it is critical to achieve greater harmony between clinicians and statisticians, and it requires considerable efforts from both sides through effective two-way communication.

A number of other issues also limited the translation of modern advanced dose finding designs and their use particularly in the pharmaceutical industry. First, the potential gain in accuracy and efficiency may be partly overshadowed by the increased complexity in study design, logistical planning and requirement of extensive simulations, which may result in potential delay in study start-up timeline. Also, because the 3+3 design is conservative and tends to underestimate the MTD, it uses fewer patients and has shorter timelines on average compared to designs like the CRM. These study duration timeline considerations may be overshadowing in some cases the fact that the 3+3 design tends to get the "wrong answer" for the estimate of the MTD more frequently than other designs. Second, although in the recent draft guidance on adaptive clinical trials, the FDA encourages the sponsors to gain experience with novel adaptive designs in exploratory studies, complex designs still have to go through rigorous regulatory reviews and it poses additional risk on the IND approval. A third issue that people often overlook is the requirement of support from statistical expertise. It is usually not a problem in large institutions and drug companies, but may be a challenge in small clinics and biotechs. Lastly, the availability of validated and user-friendly software is an important factor in adopting novel methods in order to perform model fitting in real time. All these challenges demonstrate that more work need to be done to accelerate the translation of novel designs to clinical trials.

3. Case Study: Time-to-Event Continual Reassessment Method Incorporating Cycle Information

3.1. Study Background and Method

Delayed-onset toxicities are a particular challenge for phase I trials of combination therapies (Muler et al., 2004). Most of the available dose-escalation designs require all patients have completed a fixed observation period for toxicity (eg, 1-2 cycles of the experiment regimen) before additional cohorts of patients can be enrolled. Thus, trial accrual is subject to opening and closing which may pose logistical risk on the success and completion of the study. In addition, patients who are either lost to follow-up or die of events unrelated to treatment are usually required to be replaced. Due to these reasons, the trial duration could be unacceptably long in case of prolonged observation window and unexpected high rate of patient drop-out.

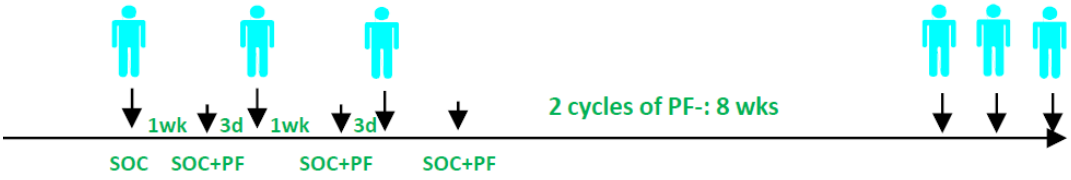
PF-566 is a fully humanized monoclonal antibody that binds to human 4-1BB. A FIH study is ongoing in combination with the standard of care (SOC) rituximab for patients with Non-Hodgkin's Lymphoma (NHL). PF-566 was administered intravenously at escalating doses of 0.03 (d_1), 0.06 (d_2), 0.12 (d_3), 0.18 (d_4), 0.24 (d_5), and 0.30 (d_6) mg/kg given once every four weeks in combination with rituximab, which was administered at a fixed dose of 375 mg/m^2, once per week for four weeks total. After evaluating the biologic mechanism and data in the preclinical PK and toxicology studies, in the phase I design the dose limiting toxicity (DLT) observation window was set as 2 cycles (8 weeks) to estimate the Maximum Tolerated Dose (MTD). However, utilizing the traditional 3+3 design or some alternative novel designs such as the CRM, the trial duration could be unacceptably long due to prolonged observation period, required enrollment gap between adjacent patients within each cohort and potential patient early drop-out (Figure 1). Furthermore, trial accrual is subject to opening and closing which may pose additional risk to the success of the study.

After reviewing different design options, the study team was finally convinced to adopt a novel statistical methodology: Time-to-Event Continual Reassessment Method (TITE-CRM), initially proposed by Cheung and Chappell (2000). It belongs to the class of Bayesian dose-finding model-based design. On top of the conventional CRM method, TITE-CRM employs a weight function (conditional probability of having a DLT during follow-up given it does occur within the DLT window) in the likelihood function. Unlike most dose finding methods, it is open to accrual continually, and makes timely dosing decision based on all the collected data from all treated patients up to that time and prior toxicity distribution.

Figure 1. Treatment Schedule for PF-05082566+SOC

DLT evaluation window: two cycles of PF-566 (2 x 4 weeks)
- SOC agent: Qweek x 4, cycle 1 only
- PF- 24 hrs after 2nd weekly dose of Standard of Care (SOC)

Minimum 1.5 week window between the first PF dose of patient 1 and the first PF dose of patient 2



Dosing schedule with conventional methods

Let T_0 be the fixed DLT evaluation time (8 weeks), T_i and x_i be the time to DLT and dose level of patient i ($i=1, \ldots, m$) respectively. $Yi(u) = I_{\{Ti \le T0 \& Ti \le u\}}$ is the DLT indicator (whether the patient has a DLT or not) for patient i at observation time u ($0 \le u \le T_0$). Denote $\Pr(T_i \le u) = \Pr(T_i \le u \mid T_i \le T_0)\Pr(T_i \le T_0) \equiv w(u, T_0)F(x_i, \theta)$, where $w(u, T_0)$ is the weight function and $F(x_i, \theta)$ denotes the dose-toxicity model. Then the weighted binomial likelihood is

$$\prod_{i=1}^{m} \{w_i F(x_i, \theta)\}^{Y_i} \{1 - w_i F(x_i, \theta)\}^{1-Y_i}$$

A common weight function is the linear (uniform) weight $w(u, T_0) = u / T_0$. We define the MTD to be the highest dose that is associated with a DLT rate $\le 25\%$. A power function modeling DLT rate at each dose d_i ($i=1, \ldots, 6$) expressed as $\Pr(DLT \mid d_i) = F(d_i, \theta)$ is used: $F(d_i, \theta) = p_i^{\exp(\theta)}$, where p_i is the prior estimate of DLT rate at dose level d_i, and $p_1 \le p_2 \le \ldots \le p_6$. These estimates were derived using the technique in Lee and Cheung (2009) to ensure model sensitivity, and when the trial is large enough the model will eventually choose a dose with DLT rate in the interval of (0.16, 0.33). θ is an unknown single parameter modeling the dose-toxicity relationship, with prior distribution

$N(0,\sigma_0^2)$, where σ_0 is the standard deviation of the normal prior distribution with mean=0. At the beginning of the trial, the initial prior value of θ is set as 0, the prior mean, which gives a prior dose-toxicity model of $F(d_i,\theta) = p_i$ based on the power function.

In the trial conduct, the first three patients will be treated at the starting dose. For each subsequent patient eligible for enrollment, the probability of DLT is estimated for each level based on all the collected data from all treated patients up to that time and the prior expectations of toxicity, and the patient is assigned to the currently estimated MTD, defined as the dose having an estimated probability of DLT closest to but not greater than the target rate (25%). The probabilities of toxicity could be estimated based on the Bayesian power model with prior distribution of the parameter to learn about the overall dose-toxicity relationship. Patients' DLT data is reported in real time to the study statistician who would estimate the MTD before the next enrolled patient is treated.

Should safety data suggest different weight (toxicity) patterns in Cycle 1 and Cycle 2 of PF-05082566, an adaptive weight function is defined in the following procedure. As a generalization, suppose the DLT evaluation window consists of *K* cycles (C_1, …, C_K), each having duration T. For patient *i* (*i=1,…,m*) with DLT, Let $Z_i=(z_{i1},\ldots,z_{iK})$ indicate the toxicity status at each cycle. Z_i ~ Multinomial(P, 1), where $P=(p_1,\ldots,p_K)$ are the probabilities of DLT occurring in each cycle

Assume P ~ Dirichlet(α_0), $\alpha_0=(\alpha_{01},\ldots,\alpha_{0K})$. Then

$$P \mid Z \sim \text{Dir}(\alpha_{01} + \sum_{i\in D_m}^{m} z_{i1},\ldots,\alpha_{0K} + \sum_{i\in D_m}^{m} z_{iK})$$

Where D_m is the set of patients with DLTs. Note that when there is no prior knowledge on time-to-toxicity distribution, $\alpha_{01}=\ldots=\alpha_{0K}$ (e.g. =1). P is estimated as

$$\hat{P}(m) = \int_0^1 Pf(P \mid Z)dP$$

Since K=2 in this study (two-cycle window of PF-05082566), The Dirichlet distribution reduces to the Beta distribution. Assume G(t) is the within-cycle toxicity cumulative distribution function, the estimated weight can be computed

$$\hat{w}(t,\hat{P}(m)) = \begin{cases} \hat{p}_1(m)G(t) & 0 < t \le T \\ \hat{p}_1(m) + \hat{p}_2(m)G(t-T) & T < t \le 2T \end{cases}$$

In the weighted binomial likelihood, derive the posterior estimate of parameter θ as

$$\hat{\theta}_m = \int \theta f(\theta \mid Y, Z, w(\hat{P}(m))d\theta$$

The $(m+1)_{th}$ patient receives the model-based MTD estimate for target tox rate *v* (=0.25)

$$x_{m+1} = \underset{d_i : F(d_i, \hat{\theta}_m) \le v}{\arg\min} \left| F(d_i, \hat{\theta}_m) - v \right|$$

To avoid overly rapid escalation and to retain the efficiency of dose administration when enrollment is fast, some restrictions are applied including no dose skipping to untested levels and every patient should be followed for at least 3 weeks before escalation occurs. There should be a minimum of 3 patients at each dose to account for interpatient variability in PK/PD data.

Dose escalation stops if any of the conditions is met: (1) the maximum sample size has been reached (n=36), (2) at least 9 patients have been treated at a dose that is predicted to be the MTD or (3) all doses appear to be overly toxic and the MTD cannot be determined in the current trial.

The study is currently ongoing and is expected to complete in 2014. Substantial savings in time thanks to the time-to-event method has already been observed. The primary trial data will be reported at a future oncology conference.

3.2. Simulations

Simulations in published literature on TITE-CRM (Cheung et al., 2000; Normolle et al., 2006; Elkind et al., 2008) and published results of Phase I clinical trials in oncology utilizing the TITE-CRM (Muler et al., 2004; Desai et al., 2007; Andre et al., 2010; Ruan et al., 2010) showed the TITE-CRM procedure dramatically shortens the trial, provides more precise and statistically sound estimates of the MTD, and most importantly, does not expose patients to significant excess risk.

To assess the operating characteristics of the design, 2000 simulations were conducted for each of the 6 hypothetical scenarios (Table 1) comparing the traditional 3+3 design with the TITE-CRM design. The true underlying dose-toxicity model as calibrated in Section 3.1 is Scenario 5. Three simulated cycle-toxicity patterns (for the 2-cycle DLT evalulation) include: a) $p_1=p_2=1/2$. b) $p_1=1/3$, $p_2=2/3$. c) $p_1=1/5$, $p_2=4/5$. The prior cycle pattern is set up as p1=p2=0.5 ($\alpha_{01}= \alpha_{02}$). Uniform time-to-event distribution is assumed within each cycle. We also assume the enrollment rate is 1 patient per week and there is a 3-week window between cohorts (n=3). The operating characteristics to be evaluated include:

- Accuracy (MTD and one level below the MTD)
- Dose allocation
- Over-dose control and safety

Figure 2 shows the operating characteristics and summary statistics of the proposed TITE-CRM design based on 3 cycle-toxicity patterns compared with the 3+3 design in each of the 6 scenarios.

The proposed design selects the correct MTD with a probability of more than 50% in all the scenarios with a maximum of 36 patients, higher than those using the standard 3+3 method (except Scenario 2). The probability of selecting the MTD and next lower dose (MTD-1) is above 90% in most scenarios and higher than that with the 3+3 in each scenario. The TITE-CRM design has a higher probability of treating patients at the MTD

(or at the MTD and MTD-1) than the 3+3 design (except for Scenario 2 where the 3+3 design treats a higher percentage of patients at the MTD with a small margin).

The chance of selecting a dose with DLT rate > 33% (toxic dose) is no more than 12% for TITE-CRM across all scenarios, and is consistently lower than that using the 3+3. The average DLT rate across simulations is well below the target 25% using TITE-CRM, even in the most toxic Scenarios 5-6. The average simulated DLT rates are similar comparing different designs in each scenario. This demonstrates that higher precision in dose estimation by using the TITE-CRM is achieved without inflating the average observed toxicity rates relative to the 3+3 design. This could be partly due to the fact the MTD is defined in a conservative manner and also due to the conservative prior setup.

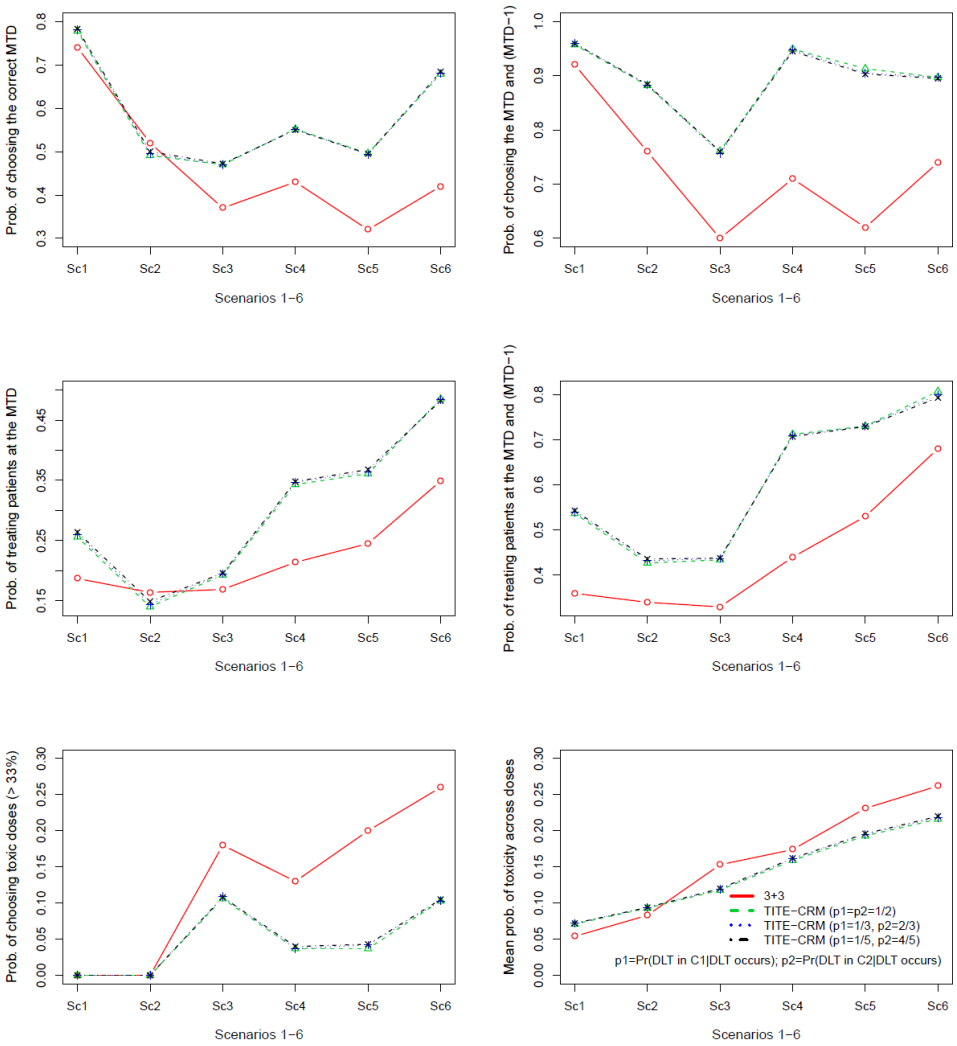
Table 1. Simulation Scenarios of the Dose-Toxicity Curve. Numbers represent the simulated DLT rates

	d_1	d_2	d_3	d_4	d_5	d_6
Scenario 1	0.01	0.02	0.02	0.04	0.07	0.15
Scenario 2	0.01	0.02	0.05	0.08	0.12	0.20
Scenario 3	0.05	0.08	0.10	0.10	0.20	0.40
Scenario 4	0.01	0.05	0.12	0.22	0.45	0.70
Scenario 5	0.08	0.15	0.25	0.40	0.55	0.70
Scenario 6	0.09	0.22	0.40	0.45	0.45	0.50

As a caveat, some published work (Normolle et al., 2006; Braun, 2006; Bekele et al., 2008) indicate that the improved operating characteristics are reduced when the weight function for a time-to-event approach significantly deviates from the true one. In particular, if DLTs are most likely to occur at the end of the DLT observation window and the weight function is set as uniform or left skewed, more patients may be exposed to overly toxic doses above the MTD when patient accrual is fast. This is because, with fast accrual, new patients are more likely to be assigned to doses considered safe but later found to be unsafe with the occurrence of late-onset toxicities (Bekele et al., 2008). We employ an adaptive weight function which is continuously updated on the basis of the time to toxicity (not necessarily DLT) distribution. Besides, at least two patients should have been on treatment for 6 weeks without DLT before escalating to the next higher dose. Thus dose escalation decision will be made more prudently.

In summary, these simulation results assist us in calibrating the method, assessing the sensitivity of the statistical model, and ensure the observed toxicity rate is within acceptable range (<25%) and patients are not exposed to excess risk based on the proposed model parameter setting and design setting. The improvement in the precision of dose finding over the standard method, coupled with the significant drop in trial duration, bode well for the operating performance of trial and justify the extra complexity in design and trial conduct.

Figure 2. Simulation Results Comparing TITE-CRM with 3+3



4. Operational Considerations in Trial Conduct

A successful dose-finding trial entails not only an adequate and efficient design, but good clinical practice in trial conduct, quality control and data collection. The statistician and the clinician are responsible for the study design, but it needs collective efforts from the entire study team (clinical, statistics, study management, data management, clinical pharmacology, drug supply etc) to ensure the trial can be conducted as designed and quality data can be collected in a timely manner.

4.1. Dose Escalation Steering Committee and Trial Conduct Process

Quality data collection is particularly important for model-based dose finding designs. This is because the dosing algorithms cannot be determined prior to study start and accurate data must be collected in time in order to carry out model fitting to determine the next level of dose for subsequent patients. In the case study, a Dose Escalation Steering Committee (DESC) advisory group was established. The DESC periodically reviews the

accumulating safety data, specifically, the DLTs. The guidance for dosing and enrollment decisions is based on the Bayesian statistical models of TITE-CRM. Other considerations may include lower grade AEs, nature and timing of the AEs, existing PK/PD data. Following each review, the DESC informs the study team and participating sites on dosing and enrollment decision.

For the trial conduct process for decision making in a model-based design as in the case study, safety data including DLTs and follow-time on treatment are collected in real time on a weekly basis and reported to the clinician and the statistician. The statistician updates the dose-toxicity model when new data come in and estimates the MTD. Prior to each DESC meeting, the statistician prepares summary slides with statistics and graphs and sends to the team for review. At the DESC meeting, the statistician presents the data and discusses dose recommendation with the rest of the committee members. Scenarios may be created to predict future dosing decisions in a foreseeable timeframe. Once a decision is made by the committee, the information will be communicated to the study team. The clinician and the study manager will inform the participating sites of the recommended dose and time of start of treatment for the next enrolled patient. In addition, patient screening may not necessarily have to wait for the decision by the DESC because it typically takes 7-10 days from screening to the start of treatment following the standard administrative procedure and a 1-week lead-in of rituximab is required in this trial. The predictive scenarios for future dosing laid out at the DESC meetings will be referred to for operational planning (e.g. communication with the site and drug supply).

Meeting minutes or summary notes of meetings, including members participating, what data are reviewed, and dose recommendation, are prepared shortly after each meeting. Each member will review the documentation and provide any edits to the statistician. The documentation should be maintained at a central location so that decision makings are reproducible.

Due to the anticipated delay in data reporting in the project database, data provided to the DESC should come from both the study database and site communication that is continually updated or revised as new information becomes available. The goal is to make the best informed decision with all existing data for dose recommendation for the subsequent patients. Any future data correction and revision should be incorporated into the statistical model accordingly.

5. Discussion

Despite the tremendous work and efforts by clinical researchers in designing practical dose finding designs that are more advanced, powerful, efficient and precise than the standard 3+3 method to accelerate improvement in cancer patients' survival and quality of life, there remains a huge lag in implementing these innovative designs in clinical trials. In this article, we present some perspectives and share two recent case studies in our hope to make a contribution to moving this field forward.

It is not surprising to see an unconscious bias regarding the 3+3 design: a method that is familiar and has been so widely used and seemly hasn't caused any serious problems or failures must be a good method. However, without readily available simulation software and modern computing power, there was no way to have known the generally poor operating characteristics of the 3+3 design and this along with the continued reluctance to

move to more principled designs may historically have contributed to the high oncology phase III failure rate . In more recent times, numerous publications in statistical and oncology journals have shown the 3+3 design exhibits undesirable statistical and clinical properties and is not as safe as many people have thought.

As clinical researchers, it is our responsibility not only to develop new and better designs, but to shepherd new methods into clinical practice. As stated in Rogatko et al. (2007), decades of experience regarding technologic innovation indicate that direct contact is the best mechanism for knowledge transfer and to reduce impediments to change current clinical research practices. Effective two-way communication between biostatisticians and clinical colleagues are required for improved uptake of alternative experimental designs in clinical trials with the ultimate beneficiary being the cancer patients.

We believe the pharmaceutical industry and academic institutions should design easy-to-use toolkits for their organization and ideally for those outside their organization to use, to streamline the design and planning processes to minimize the extra time assumed compared to when using a standard approach. For example, MD-Anderson cancer center has done excellent work and developed free easy-to-use software on their website for oncology clinical trials.

Also importantly, regulatory agencies who are the gate-keepers of drug development and clinical trial designs should proactively encourage the adoption and implementation of novel statistical designs in clinical practice. As mentioned in Section 2.4, one of the major obstacles in convincing the clinical team to use these new methods is the fear of a longer review time and increased risk of rejection of the investigational new drug submission. The fear is unfounded but commonly seen. It is another indicator that lack of communication between the regulatory agency and the sponsor is an impediment to innovation. After years of joint efforts from the industry and regulatory agency, some progress has been made. The FDA released a draft guidance document on adaptive clinical trials in February 2010, which generally is welcomed by the industry. In particular, in the draft document FDA emphasized the natural role of adaptive designs in exploratory studies, and “encourages sponsor to gain experience with these adaptive design methods in this setting”.

In summary, it has been shown that the 3+3 design has generally poor properties and leads often to an incorrect estimate of the MTD which may ultimately increase the risk of a negative phase III trial and reduce the likelihood of a compound getting approved denying its availability to cancer patients. Like in most phase II and phase III studies, the study team should explore thoroughly different design options in phase I and select the optimal dose-finding approach for their study. The time invested will be paid back with an enhanced clinical trial design and improved probability of technical and regulatory success in the long run.

References

1. Storer BE. Design and analysis of Phase I clinical trials. *Biometrics*. 1989; **45**:925–937.
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.

3. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A. Translation of innovative designs into phase I trials. *Journal of Clinical Oncology*. 2007; *25*(31), 4982-4986.
4. Lin Y, Shih WJ. Statistical properties of the traditional algorithm-based designs for phase I cancer trials. *Biostatistics*. 2001; 2: 203-215
5. Ivanova A. Escalation, group and A+B designs for dose-finding trials. *Statistics in Medicine*. 2006; 25: 3668-3678.
6. Durham SD, Flournoy N, Rosenberger WF. A random walk rule for phase I clinical trials. *Biometrics*. 1997; 53: 745-760.
7. Simon R, Rubinstein L, Arbuck SG, Christian M. Accelerated titration designs for phase I clinical trials in oncology. *JNCI.* 1997; 89: 1138-1147.
8. Leung DHY and Wang YG. Isotonic designs for phase I trials. *Controlled Clinical Trials*. 2001; 22: 126-138.
9. Oron AP, Hoff P. Small-sample behavior of novel phase I cancer trial designs. *Clinical Trials*. 2013; 10(1), 63-80.
10. Ji Y, Liu P, Li Y, Bekele B. A modified toxicity probability interval method for dose-finding trials. *Clinical Trials*. 2010; *7*(6), 653-663.
11. Liu S, Yuan Y. Bayesian Optimal Interval Designs for Phase I Clinical Trials. *Statistics in Medicine*, 2012; revision invited.
12. 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.
13. Faries D. Practical modifications of the continual reassessment method for Phase I cancer clinical trials. *Journal of Biopharmaceutical Statistics* 1994; **4**:147–164.
14. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**:673–684.
15. Cheung YK. Dose Finding by the Continual Reassessment Method. Taylor & Francis US. 2011.
16. Shen L, O'Quigley J. Consistency of continual reassessment method under model misspecification. *Biometrika*. 1996; *83*(2), 395-405.
17. Cheung Y. Coherence principles in dose-finding studies. *Biometrika*. 2005; *92*(4), 863-873.
18. Huang B, Chappell R. Three-dose–cohort designs in cancer phase I trials. *Statistics in medicine*. 2008; *27*(12): 2070-2093.
19. Braun T. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Controlled Clinical Trials*. 2002; *23*(3), 240-256.
20. Thall P, Cook J. Dose-Finding Based on Efficacy–Toxicity Trade-Offs. *Biometrics*. 2004; *60*(3), 684-693.
21. Fox E, Curt G, Balis F. Clinical trial design for target-based therapy. *The oncologist*. 2002; *7*(5), 401-409.
22. Piantadosi S, Liu G. Improved designs for dose escalation studies using pharmacokinetic measurements. *Statistics in medicine*. 1996; *15*(15), 1605-1618.
23. Mandrekar S, Cui Y, Sargent D. An adaptive phase I design for identifying a biologically optimal dose for dual agent drug combinations. *Statistics in medicine*. 2007; *26*(11), 2317-2330.
24. Polley M, Cheung Y. Two-Stage Designs for Dose-Finding Trials with a Biologic Endpoint Using Stepwise Tests. *Biometrics*. 2008; *64*(1), 232-241.
25. Dancey J, Chen H. Strategies for optimizing combinations of molecularly targeted anticancer agents. *Nature reviews Drug discovery*. 2006; *5*(8), 649-659.
26. Le Tourneau C, Lee J, Siu L. Dose escalation methods in phase I cancer clinical trials. *Journal of the National Cancer Institute*. 2009; *101*(10), 708-720.

27. Harrington J, Wheeler G, Sweeting M, Mander A, Jodrell D. Adaptive designs for dual-agent phase I dose-escalation studies. *Nature Reviews Clinical Oncology*. 2013; *10*(5), 277-288.
28. Kramar A, Lebecq A, Candalh E. Continual reassessment methods in phase I trials of the combination of two drugs in oncology. *Statistics in medicine*. 1999; *18*(14), 1849-1864.
29. Thall P, Millikan R, Mueller P, Lee S. Dose-Finding with Two Agents in Phase I Oncology Trials. *Biometrics*. 2003; *59*(3), 487-496.
30. Huang X, Biswas S, Oki Y, Issa J, Berry D. A parallel phase I/II clinical trial design for combination therapies. *Biometrics*. 2007; *63*(2), 429-436.
31. Yuan Y, Yin G. Sequential continual reassessment method for two-dimensional dose finding. *Statistics in medicine*. 2008; *27*(27), 5664-5678.
32. Yin G, Yuan Y. A latent contingency table approach to dose finding for combinations of two agents. *Biometrics*. 2009; *65*(3), 866-875.
33. Bailey R. Designs for dose–escalation trials with quantitative responses. *Statistics in Medicine*. 2009; 28: 3721–3738.
34. O'Quigley J. Commentary on "Designs for dose-escalation trials with quantitative responses". *Statistics in Medicine*. 2009; 28: 3745-3750.
35. Cheung YK, Chappell R. Sequential designs for phase 1 clinical trials with late-onset toxicities. Biometrics. 2000; 56: 1177-82.
36. Lee S, Cheung YK. Model calibration in the continual reassessment method. *Clinical Trials.* 2009; 6: 227-238.
37. Bekele BN, Ji Y, Shen Y, Thall PF. Monitoring late-onset toxicities in phase I trials using predictive risks. *Biostatistics*. 2008; 9: 442-57.
38. Braun TM. Generalizing the TITE-CRM to adapt for early and late onset toxicities. *Statistics in Medicine*. 2006; 25: 2071-83.
39. Desai SP, Ben-Josef E., Normolle D et al. Phase I study of Oxaliplatin, full-dose gemcitabine, and concurrent radiation therapy in pancreatic cancer. *J Clin Oncol.* 2007; 25: 4587-92.
40. Muler J, McGinn C, Normolle D et al. Phase I trial using a time-to-event continual reassessment strategy for dose escalation of cisplatin combined with gemcitabine and radiation therapy in pancreatic cancer. *J Clin Oncology*. 2004; 22: 238-243.
41. Normolle D, Lawrence T: Designing dose-escalation trials with late-onset toxicities using the time-to-event continual reassessment method. *J Clin Oncol.* 2006; 24: 4426-4433.
42. Ji Y, Wang SJ. Modified Toxicity Probability Interval Design: A Safer and More Reliable Method Than the 3+ 3 Design for Practical Phase I Trials. *Journal of Clinical Oncology*. 2013; 31(14), 1785-1791.
43. Fanale M et al. Phase I study of bortezomib plus ICE (BICE) for the treatment of relapsed/refractory Hodgkin lymphoma. *British Journal of Haematology*. 2011; 154:284 286.
44. Dykstra R, Robertson T. An algorithm for isotonic regression for two or more independent variables. *Ann Stat*. 1982; 10:708-716.
45. Ting N. General Considerations in Dose-Response Study Designs. *Dose Finding in Drug Development*. 2006; Springer, New York pp 1-17.
46. Harrington J, Wheeler G, Sweeting M et al. Adaptive designs for dual-agent phase I dose-escalation studies. *Nature Reviews Clin Onc.* 2013; 10: 277-288.