Original Research Paper

ATLCEP: A New Integrated Design that Encompasses the Safety and Efficacy Objectives of Phase I/II Trials

[1]**Revathi Ananthakrishnan,** [2]**Stephanie Green,** [1]**Mark Chang,**
[1]**Gheorghe Doros,** [1]**Joseph Massaro,** [3]**Daniel Li and** [1]**Michael LaValley**

[1]*Department of Biostatistics, Boston University, 801 Massachusetts Avenue 3rd Floor, Boston, MA-02118, USA*
[2]*New London, CT-06320, USA*
[3]*Juno Therapeutics, Seattle, WA, USA*

Corresponding Author:
Revathi Ananthakrishnan
Department of Biostatistics,
Boston University, 801
Massachusetts Avenue 3rd
Floor, Boston, MA-02118, USA
Email: revathia@gmail.com

Abstract: Several designs, such as designs to find the optimal biological dose, the Eff-Tox design and seamless Phase I/II designs, have been proposed to evaluate both drug toxicity and efficacy as alternatives to the traditional paradigm of a stepwise drug development approach. Here, we first examine the effect of sample and cohort size on the accuracy of dose selection in early phase oncology designs and then propose a design that is large enough to allow accurate dose selection for toxicity and that incorporates Bayesian decision rules at the end to select an optimal dose for toxicity and efficacy. We propose the Accelerated Titration Large Cohort Early Phase (ATLCEP) design, a moderately large, simple rule-based integrated Phase I/II trial design that evaluates both safety and efficacy. This design incorporates stopping rules within dose levels to allow more flexible decision-making. Finally, we compare the operating characteristics of this design with other Phase I/II strategies, via simulations. Our simulations of the ATLCEP design yield a mean sample size of approximately 42 patients for the true DLT and response rates and stopping rules considered and show that with this sample size the design can robustly pick a dose that is optimal for efficacy and safety. In our simulations, it performs as well as or better than the Eff-Tox or the Optimal Biological Dose (OBD) Isotonic design. It also performs better than a 3+3 Phase I design followed by a standard Phase II design.

Keywords: Phase I/II Integrated Design, Rule-Based Design, Eff-Tox Design, OBD Design

Introduction

In a Phase I oncology clinical trial, the safety of the investigational drug is studied and the Maximum Tolerated Dose (MTD) to be used in a Phase II trial is determined (Huang *et al*., 2016; Yang *et al*., 2015). In a Phase II trial, in addition to safety, the efficacy of the drug is investigated. The number of patients enrolled in a Phase I oncology trial is usually very small with 15-30 patients, while the number of patients enrolled in a Phase II trial is larger with 50-100 patients.

With such a small number of patients, Phase I dose-finding oncology trials do not always accurately select the MTD. The 3+3 design that is still frequently used in Phase I dose-finding oncology trials has been shown to

be inaccurate in determining the MTD and to under-dose many patients (Hansen *et al*., 2014). Furthermore, the efficacy of novel anti-cancer agents may not always increase with an increase in dose and can peak at any dose level (Sato *et al*., 2016; Wages and Tait, 2015). Thus, Phase I trials may not target the optimal dose taking into consideration both toxicity and response. Assessment of both Dose Limiting Toxicities (DLTs) and efficacy responses in a larger trial has the potential for a more accurate determination of a suitable dose, compared to a Phase I followed by a Phase II trial.

The dose that is commonly used in a Phase III oncology trial is the MTD determined from an earlier dose-finding Phase I trial, as Phase II oncology trials do not typically evaluate multiple doses. If the MTD obtained from the

Phase I trial is not correct, this can have enormous cost and resource repercussions for the development of an oncology drug. If the dose administered to patients in the Phase III trial is too high, the trial can fail because of many early withdrawals due to toxicity. On the other hand, if the dose administered to patients is not efficacious, it may not be possible to determine from the Phase III trial results whether the drug did not work at all or whether just the dosage was wrong. Thus, conducting small early phase trials and rapidly moving on to the corresponding Phase III trial may not be a good strategy and it may be worthwhile to spend more time and treat more patients in the early phase trials to thoroughly investigate and accurately determine the optimal dose for safety and efficacy.

The importance of selecting the right safe dose and the optimal dose for toxicity and efficacy has been illustrated several times. In a paper by Markman (2010), the drug pegylated liposomal Doxorubicin (PLD), used to treat platinum resistant ovarian cancer and its dosage are discussed. The dose of PLD explored in initial clinical trials and approved by the FDA was 50 mg/m^2 administered every 28 days. However, clinical experience has shown that this dose leads to substantial numbers of adverse effects, while a lower dose of 40 mg/m^2 is equally efficacious but leads to fewer adverse effects, as emphasized by Ferrandina (2010). For other examples, a paper by Schilsky *et al*. (2013) provides a summary of several oncology drugs and states that for many of these, the dose that is clinically administered is often lower, or higher, than the dosage that was approved by the regulatory authority.

In this study, we propose a simple integrated Phase I/II trial design called the "*Accelerated Titration Large Cohort Early Phase (ATLCEP) design*" that evaluates both safety and efficacy with a moderately large sample size. The motivation for exploring a larger rule-based design comes from the substantial effect of sample size and/or cohort size on the accuracy of dose selection in early phase oncology designs. With 35-50 patients, the safety of the drug and the MTD can be evaluated with greater accuracy than with 15-30 patients and the efficacy of the study drug can also be investigated in this larger sample. Thus, a single, larger study could serve to determine a dose that is optimal for both safety and efficacy. Several designs have been proposed to evaluate both drug toxicity and efficacy in early phase trials such as designs to find the optimal biological dose (Wages and Tait, 2015; Zang *et al*., 2014) and the Eff-Tox design (Thall and Cook, 2004; Thall *et al*., 2006) among others (Pan *et al*., 2013; 2014; Hoering *et al*., 2011; 2013; Braun, 2002; Yin *et al*., 2006; Zhang *et al*., 2006; Dragalin *et al*., 2008; Li *et al*., 2017); a recent reference by Yuan covers Bayesian designs for Phase I/II trials (Yuan *et al*., 2016).

Our proposed rule-based ATLCEP design is similar in concept to the Eff-Tox design. The Bayesian decision

rules incorporated in our design at the end of the trial for selecting an optimal dose for efficacy and safety are the same as those used in the Eff-Tox design for determining the dose level to which the next cohort of patients should be assigned, based on the number of DLTs and responses in previous patients. However, our design does not require the challenging selection of three points to define the trade-off function contour used to determine the optimal dose for the next cohort of patients in the Eff-Tox design. In addition, unlike these other designs, the proposed rule-based design is easily implemented without the assistance of a statistician during the trial conduct.

Methods

Effect of Sample Size/Cohort Size on the Accuracy of MTD or Optimal Dose Selection

We first simulate the effect of sample and/or cohort sizes on the accuracy of MTD selection in rule-based designs such as the 3+3, 5+5a, 10+10 and 20+20 designs as well as the Continual Reassessment Method (CRM) design. For this, we use a logistic dose-toxicity curve, which is often employed to describe the relation between dose and toxicity (Garrett-Mayer, 2006); we also study the effect of sample size and cohort size on the accuracy of optimal dose selection in the Eff-Tox design (Thall *et al*., 2006) [see 'Simulations' Section for details]. We compare the results for accuracy of optimal dose selection of the ATLCEP design to those of the Eff-Tox design and the OBD Isotonic design for various scenarios of true toxicity and efficacy rates. We also compare the results of the ATLCEP design to those of a 3+3 Phase I design followed by a Phase II design. We briefly describe each design studied in this study below.

Design Descriptions

The CRM design is a Bayesian design, where the cumulative DLT data along with a pre-specified dose-toxicity model, frequently a one or two parameter logistic model, is used to assign the next patient(s) to a dose level (O'Quigley *et al*., 1990; Garrett-Mayer, 2006). The stopping point of the trial is usually when the pre-specified sample size is reached.

The Eff-Tox design is also a Bayesian design, considering the trade-off between the probabilities of drug toxicity and efficacy to determine the optimal dose for each new cohort of patients. This requires the selection of three points to define a trade-off function contour, which needs careful consideration. If the trade-off contour is not sufficiently steep, the design can get stuck at a low dose and fail to find an optimal dose for the assignment of the next cohort of patients (https://biostatistics.mdanderson.org/softwaredownload/ProductSupportFiles/EffTox/EffToxUsersGuide.pdf).

The stopping point of the trial is usually when the pre-specified sample size is reached.

The details of the OBD Isotonic design are provided by Zang *et al.* (2014). To determine the OBD an admissible set of doses, which satisfy a safety criterion similar to that used in the Eff-Tox design, is first defined. The OBD is then the lowest dose with the highest response rate within the admissible set. The stopping point of the trial is usually when the pre-specified sample size is reached.

The 3+3 and 5+5a rule based designs target a DLT rate of approximately 0.2 (Table 4.1 of Chapter 4 of Ting (Ivanova, 2006b) and Table 1 of Ananthakrishnan *et al.* (2017). The basic procedure for these designs are presented below:

- 3+3 Design: Enroll 3 patients at the lowest dose level

  - If 0 out of 3 enrolled patients have a DLT, then escalate to the next dose level and enroll 3 more
  - If 1 out of 3 patients has a DLT, then add 3 more patients at the same dose level; if 2 or more patients out of 3 or 6 patients experience a DLT, then stop the trial
  - The MTD is one dose level below the last dose level examined

- 5+5a Design: Enroll 5 patients at the lowest dose level

  - If 0 out of 5 enrolled patients have a DLT, then escalate to the next dose level and enroll 5 more
  - If 1 or 2 out of 5 patients have a DLT, then add 5 more patients at the same dose level; if 3 or more patients out of 5 or 10 patients experience a DLT, then stop the trial
  - The MTD is one dose level below the last dose level examined

The 10+10 and 20+20 designs we present are constructed so that they target a DLT rate of approximately 0.2, (see the Appendix for the target DLT interval of the 20+20 design):

- 10+10 design: Enroll 10 patients at the lowest dose level

  - If <=2 patients out of 10 have a DLT, then enroll 10 patents in the next higher dose level
  - If 3 or 4 patients experience a DLT, then enroll 10 more patients in the same dose level
  - If <=4 patients out of 20 have a DLT, then enroll 10 patients in the next higher dose level
  - If 5 or more patients out of 10 or 20 experience a DLT at a dose level, then stop the trial
  - The MTD is one dose level below the last dose level examined

- 20+20 design: Enroll 20 patients at the lowest dose level

  - If <= 6 patients out of 20 have a DLT, then enroll 20 patents in the next higher dose level
  - If 7 or 8 patients experience a DLT, then enroll 20 more patients in the same dose level
  - If <= 8 patients out of 40 have a DLT, then enroll 20 patients in the next higher dose level
  - If 9 or more patients out of 20 or 40 patients have a DLT at a dose level, then stop the trial
  - The MTD is one dose level below the last dose level examined

Proposed Design

We propose the *ATLCEP* design that is a simple rule-based design during the conduct of the trial, which enrolls larger sample sizes than standard Phase I designs and incorporates Bayesian decision rules for optimal dose selection at the end of the trial. The idea is to use accelerated titration to avoid widespread underdosing of patients, but to switch to the large cohort phase near the MTD to gain accuracy.

The schematic of the ATLCEP design is shown in Figure 1 and the design is described here. The ATLCEP design starts in the accelerated titration phase by enrolling patients in cohorts of size 3:

- The first cohort of 3 patients is assigned to the lowest dose level and subsequent cohorts of size 3 continue to be assigned to increasing dose levels as long as none of the 3 patient cohorts experience a DLT
- When one or more patients in a cohort of size 3 experience a DLT, then the design switches from the accelerated titration phase into the large cohort phase which is based on a modification of the 20+20 design presented earlier

In the large cohort phase, batches of 6 or 8 patients are used to fill in the 20-patient cohorts to limit the number of patients that are exposed to the study drug at once:

- 3 new patients are enrolled at the same dose where accelerated titration stopped to reach an initial total of 6
- If ≥4 patients out of 6 have a DLT, then the trial stops
- If <4 patients out of 6 have a DLT, then 8 more patients are added to this dose.
- If >8 patients out of 14 have a DLT, then the trial stops
- If ≤ 8 patients out of 14 have a DLT, then 6 more patients are added to this dose
- If > 8 patients out of 20 have a DLT, the trial stops.
- If < 7 patients out of 20 have a DLT, the dose is escalated and 6 patients are treated at the next higher dose level and the process starts over at that higher dose level

- If 7 or 8 patients out of 20 have a DLT, then the same dose level is expanded (up to 40) and 6 additional patients are treated at the same dose
- If >8 out of 26 patients have a DLT, then the trial stops
- If ≤8 patients out of 26 have a DLT, then 8 additional patients are added to the dose
- If > 8 out of 34 patients have a DLT, then the trial stops

- If ≤ 8 out of 34 patients have a DLT, then 6 additional patients are added to the dose.
- If >8 out of 40 patients have a DLT, then the trial is stopped
- If ≤ 8 out of 40 patients have a DLT, the dose is escalated and 6 patients are treated at the next higher dose level and the process starts over at that higher dose level
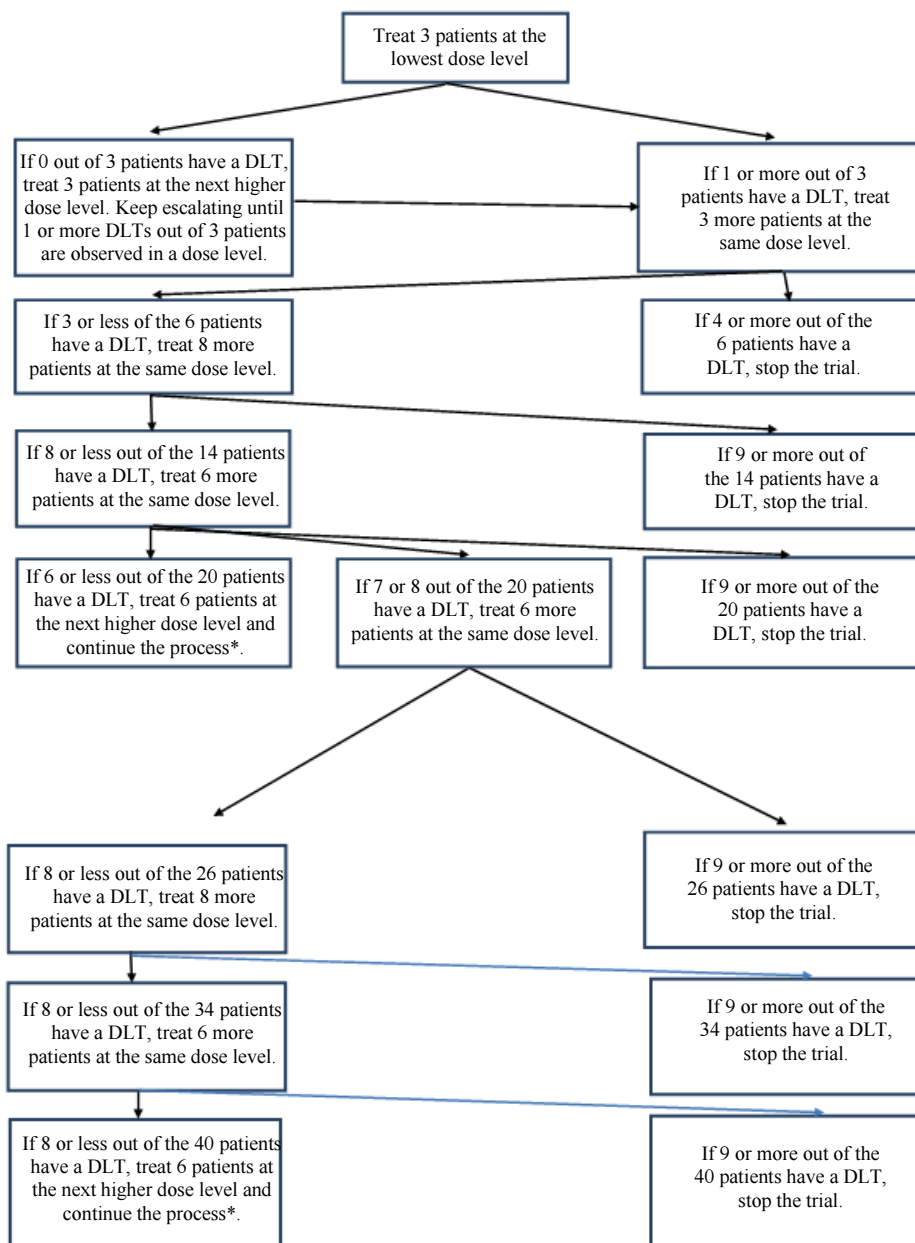

Fig. 1: Schematic of the ATLCEP design. At the end of the trial, the dose that is optimal for safety and efficacy is chosen using Bayesian decision rules and other criteria; *Note that we can escalate if we see 6 or less DLTs out of 20 patients or 8 or less DLTs out of 40 patients in a dose level, but we can also escalate if 0 DLTs and 0 responses are observed in the first 14 patients in a dose level since we cannot observe more than 6 DLTs in the last 6 patients and since the dose is not efficacious (otherwise we enroll the next 6 patients at the same dose level and continue the process)

In summary, we can escalate to the next dose level if ≤6 out of 20 patients in a dose level experience a DLT or ≤8 out of 40 patients in a dose level experience a DLT, similar to the escalation rules of the 20+20 design described earlier. However, we can also escalate if out of the first 14 patients in a dose level, we observe 0 DLTs and 0 responses, since no more than 6 DLTs can be observed in the last 6 patients and the dose is not efficacious. The stopping rules of the trial are described above. Additional patient safety measures can be implemented with a safety review committee that can stop the trial at its discretion at any point.

At the end of the trial, a dose that is acceptable for safety and efficacy is chosen using the following Bayesian decision rules on the posterior probabilities of p_i and q_i, which are the toxicity and efficacy probabilities at dose i:

$$Pr\left(p_i < p_{TP} \mid data\right) > a \text{ and } Pr\left(q_i > p_{EP} \mid data\right) > b$$

Here, p_{TP} and p_{EP} are the upper limit for toxicity and lower limit for efficacy respectively, whose values are pre-specified for the study based on discussions with clinicians and "a" and "b" are small probability cut-offs (The cutoff probabilities are typically 0.1 or smaller in value (https://biostatistics.mdanderson.org/softwaredownload/ProductSupportFiles/EffTox/EffToxUsersGuide.pdf) and we use the upper limit of 0.1 for all the simulations in this study). Here, we use $p_{TP} = 0.33$, $p_{EP} = 0.5$ and $a = b = 0.1$. We also assume that both p_i and q_i follow a Jeffreys prior i.e., a uniform Beta (0.5, 0.5) prior (Pan *et al.*, 2014). The posterior distributions of p_i and q_i are then Beta (0.5+ x_i, 0.5+n_i-x_i) where x_i is the number of DLTs or responses respectively at dose level i out of n_i patients at that dose level. These Bayesian decision rules are the same as those used in the Eff-Tox design to assign new cohorts of patients a dose and they ensure that doses that are too toxic or that are too inefficacious are not selected.

If more than one dose is found to be acceptable for safety and efficacy in a trial using the Bayesian decision rules above, then the following criteria can be used, in the suggested order of preference, to choose a single dose level: (a) the value of a pre-specified utility function evaluated at each dose level or (b) the percentage of patients who respond but do not have a DLT at each dose level or (c) an empirical Odds Ratio (OR) (http://www2.ims.nus.edu.sg/Programs/011wclinic/files/guosheng_ppt.pdf) of toxicity to response at each dose level. In this case, we would select the dose that has the maximum value for the utility function or that has the largest percentage of patients with a response but no DLT or that has the smallest value for the empirical odds ratio. We provide an example utility function for criterion (a) and show the calculation for the empirical OR for criterion (c) below.

One possible utility function would be the fraction of responders at each dose level minus a constant 'c' multiplied by the fraction of patients with DLTs at that dose level. For dose level i, the formula used is:

$$utilit\, y_i = r_i - c * d_i$$

where, c is a constant that can vary between 0 and 1, r_i is the fraction of patients having a response and d_i is the fraction of patients with a DLT. This utility function was also employed in Ivanova *et al.* (2009).

The calculation for the empirical odds ratio at a dose level is based on the numbers of subjects, the number of responses and the number of DLTs at the dose level. The following formula is used:

$$OR_i = \frac{number\ of\ observed\ DLTs\ at\ dose\ i * \left(\begin{array}{c} number\ of\ patients\ at\ dose\ i - number \\ of\ observed\ responses\ at\ dose\ i \end{array}\right)}{number\ of\ observed\ responses\ at\ dose\ i * \left(\begin{array}{c} number\ of\ patients\ at\ dose\ i - number\ of \\ observed\ DLTs\ at\ dose\ i \end{array}\right)}$$

Simulations

We simulated the performance of the proposed ATLCEP design using SAS. In many simulations to study the statistical operating characteristics of the ATLCEP design, we generate the true DLT rate at each dose level (p_i) using a logistic dose-toxicity curve, as in Table 1, because the toxicity of an anti-cancer agent typically increases with an increase in dose. The two coefficients of the logistic dose toxicity curve used in Table 1 are calculated using the following parameters: The true DLT rate at dose level 1 of 100 units is 0.01 and the true DLT rate at the MTD dose level (dose level 4 of 501 units) is 0.2. However, we select the true response rate at each dose level (q_i) manually (Table 1) because the efficacy of an anti-cancer agent may not always increase with an increase in dose and can peak at any dose level (Sato *et al.*, 2016; Wages and Tait, 2015). We generate two binary random variables X1~ Bernoulli(p) and X2~ Bernoulli(q) for toxicity and efficacy respectively, which can either be correlated or uncorrelated.

In Table 1, we select the true response rates such that the true response rate peaks at dose level 4. However, different dose-response curves can be investigated (see results in the Appendix for the ATLCEP design for other scenarios of true response rates).

Table 1: True underlying DLT and response rates

Dose	Dose level	True probability of toxicity at each dose, generated from a logistic curve*	True probability of response at each dose is selected manually
100 units	1	0.01	0.01
200	2	0.02	0.05
334	3	0.06	0.15
501	4	0.2	0.45
701.4	5	0.55	0.2
932.86	6	0.89	0.05

Log_e(DLT rate/(1-DLT rate)) = -5.39533+0.008002 dose

For the rule-based designs, we use *SAS* to simulate the binary toxicity and response and to carry out the rules for escalation and stopping to determine the effect of sample size on the accuracy of MTD selection, as described in Ananthakrishnan *et al*. (2017). In the main paper, we evaluated the ATLCEP design in 10000 simulated trials with the true DLT and response rates in Table 1, with correlation coefficient $r = 0$ (see Appendix for other scenarios for which we evaluated the design). We used a correlation coefficient r of 0 in our main simulations to be conservative, as Cai *et al*. (2014) showed that joint modeling of efficacy and safety does not necessarily improve the performance of the dose finding, especially when efficacy is weakly correlated with toxicity. Calculations with other correlation values are discussed briefly in the Results Section and the Appendix. For each simulated trial, we summarize the number of patients, the number of DLTs and the number of responses at each dose. We then use the Bayesian decision rules described, to select an acceptable dose for toxicity and efficacy at the end of each simulated trial. We also estimate the value of the utility function at each dose, the percentage of patients at each dose who respond but do not have a DLT and the empirical odds ratio at each dose, for use in optimal dose selection. The accuracy of MTD selection, the median and maximum sample sizes are simulation outputs for these rule-based designs.

For the CRM design, we use the *R* package *CRM* to study the effect of sample size and cohort size on the accuracy of MTD selection with the input parameters given in the Appendix.

For the Eff-Tox design, we use the *Eff-Tox design* package from the MD Anderson Cancer Center to study the effect of sample size and cohort size on the accuracy of dose selection with the input parameters given in the Appendix.

Results

Effect of Sample Size and Cohort Size on Dose Selection

Table 2 shows the percentage of times that the true MTD (dose level 4) is selected by various rule-based designs that allow only escalation and that target a DLT rate of approximately 0.2. To produce these results, we use the true DLT rates shown in Table 1. We perform 10000 simulations for each of the rule-based designs in Table 2. For comparison, we also include in Table 2 the results of two specific cases of the CRM design using the input parameters given in the Appendix. As can be seen, the accuracy of MTD selection increases with an increase in sample size for all the cases considered.

We also evaluated other scenarios for the CRM design than those shown in Table 2, as well as many scenarios for the Eff-Tox design (results not shown here). In these examples that we studied, we found that the accuracy of MTD or dose selection improved dramatically with an increase in sample size for all the cases and designs considered. Larger cohort sizes may result in a small reduction in the accuracy of MTD selection for the CRM design, but could improve dose selection in the Eff-Tox design. Thus, cohort size and sample size are crucial parameters to explore, in the design of an early phase oncology trial.

Simulation Results for the Accelerated Titration Large Cohort Early Phase (ATLCEP) Design

From the simulation results in Table 2, we observe that the 20+20 design has a high probability of selecting the MTD due to its large sample size. Our simulations yield a median sample size of 100 for the 20+20 design for the true DLT rates in Table 1. This larger sample size allows us to consider addressing drug efficacy in addition to drug safety. However, the sample size of the 20+20 design is relatively large for an integrated Phase I/II trial and a sample size closer to 50 would be more reasonable. Therefore, we proposed the Accelerated Titration Large Cohort Early Phase (ATLCEP) design, for which design simulations yielded a mean sample size of 42 and a median sample size of 35 for the true DLT and response rates in Table 1. Basing the large cohort phase on the 10+10 or 15+15 designs would yield even smaller mean sample sizes, but would also result in less accurate dose selection and hence was not considered further. Our results for the ATLCEP design, which considers both drug toxicity and efficacy in selecting an optimal dose, are shown in Table 3.

Table 2: Accuracy of MTD selection in rule-based designs and for the CRM design

Design	Target DLT rate	% of Times the True MTD (Dose Level 4) is Selected	Median Sample Size, Maximum Sample Size
Standard 3+3 design	0.17-0.26	**60.0%**	**15, 33**
5+5a design	0.2-0.25	**65.9%**	**30, 50**
CRM 5, 50 (cohort size of 5, sample size of 50)	0.15-0.25	**81.6%**	50 (maximum sample size, which is an input)
10+10 design	0.2-0.24	**74.0%**	**50, 80**
20+20 design	0.2-0.21	**90.1%**	**100, 140**
CRM 20,120 (cohort size 20, sample size 120)	0.15-0.25	**90%**	120 (maximum sample size, which is an input)

Table 3: Results for the ATLCEP design

	Dose Level 1	Dose Level 2	Dose Level 3	Dose Level 4	Dose Level 5	Dose Level 6
True toxicity, true efficacy rate	0.01, 0.01	0.02, 0.05	0.06, 0.15	0.2, 0.45	0.55, 0.2	0.89, 0.05
Odds of toxicity to efficacy from simulations at Dose Level I (OR_i)*	**0.20**	**0.26**	**0.35**	**0.31**	**7.71**	**20.05**
Average number of patients	**3.5**	**4.5**	**7.3**	**14.0**	**12.2**	**0.28**
Average number of DLTs	**0.04**	**0.1**	**0.45**	**2.8**	**6.7**	**0.25**
Average number of responses	**0.04**	**0.22**	**1.1**	**6.3**	**2.4**	**0.01**
Average number of patients with a Response but no DLT	**0.04**	**0.22**	**1.0**	**5.0**	**1.1**	**0.0**
Decision based on the Bayesian posterior Probabilities (% of times out of 10000 simulations each dose satisfies the Bayesian decision criteria for both toxicity and efficacy)**	**2.86%**	**13.02%**	**28.6%**	**76.08%**	**1.87%**	**0%**
% of times out of 10000 simulations each dose level is the one with the maximum value of the utility function						
when c = 0.1***	**1.54%**	**4.46%**	**11.91%**	**76.15%**	**5.93%**	**0.01%**
when c = 0.5	**4.6%**	**6.66%**	**16.05%**	**71.63%**	**1.06%**	**0%**
when c = 1	**7.34%**	**8.64%**	**20.79%**	**63.04%**	**0.19%**	**0%**

* Calculated using only those simulations runs with a non-zero denominator for OR_i. For the lower dose levels (levels 1 and 2), the denominator is zero in many simulation runs since the average number of responses is zero.

** No dose level is chosen as acceptable for toxicity and efficacy~15% of the time. The addition of the percentages for dose selection based on the Bayesian decision rules can add up to more than 100 since more than one dose level can be chosen as acceptable for toxicity and efficacy in each simulation. These numbers for dose selection can change depending on the probability cut-off values "a" and "b" used in the Bayesian decision rules for safety and efficacy but the dose selection numbers using the utility function remain the same.

***c = 1 gives equal weight to toxicity and efficacy while 0.1 gives a very small weight to toxicity and more weight to efficacy.

Mean sample size for this example is 41.75; median sample size is 35, minimum sample size is 12 and maximum sample size is 126.

Our simulation results in Table 3 for the ATLCEP design show that for the true DLT and response rates in Table 1, dose level 4 is the dose that satisfies the Bayesian decision rules for safety and efficacy most frequently. Dose level 4 is chosen as the dose level that is acceptable for safety and efficacy in 76% of the simulation runs in this case (Fig. 2). In an actual trial, if more than one dose level satisfies these two Bayesian decision rules, other criteria such as the value of the utility function, the percentage of patients with a response but no DLT, or the OR_i can be used to choose a single dose level. In our example, dose level 4 has the maximum value of the utility function most frequently and has the maximum value for the average percentage of patients with a response but no DLT. The minimum value for the OR_i from the simulations

is at dose level 1, not at dose level 4, but there are very few responders in dose level 1 (see the first footnote to Table 3). Hence, based on these results of the ATLCEP design, our final selection for optimal dose is dose level 4 for this example.

Corresponding results for the ATLCEP design with several other scenarios of true toxicity and efficacy rates are shown in the Appendix.

In general, the results in Table 3 can be investigated for different values of the correlation coefficient *r*. However, *r* can take on only certain values and the highest value that *r* can take will differ for different combinations of true DLT and response rates (see the Appendix). For a logistic dose-toxicity curve and for a monotonically increasing dose-response curve, a higher value of r can be used (see the Appendix).
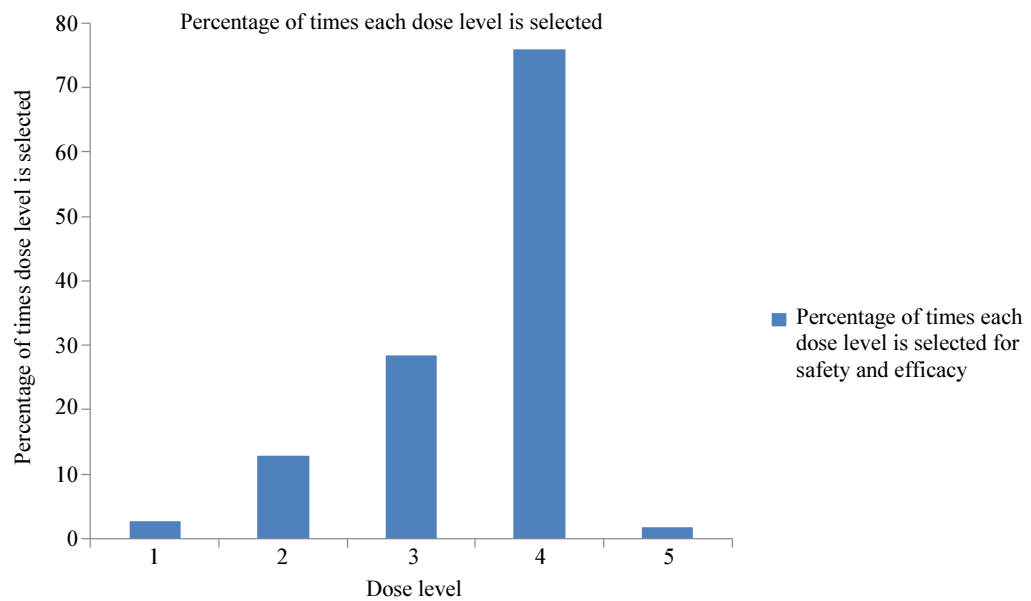
Fig. 2: Percentage of times out of 10000 simulations that each dose level is selected as acceptable for safety and efficacy in the ATLCEP design, based on the true toxicity and response rates in Table 1

Table 4: Comparing the Results of the Eff-Tox design and the OBD Isotonic Design to those of the ATLCEP Design for Various Scenarios of True Toxicity and Efficacy Rates

Scenario		Dose level					
		1	2	3	4	5	None
1	True toxicity, true efficacy rate	0.05, 0.10	0.10, 0.30	0.15, 0.60	0.30, 0.62	0.45, 0.65	-
	Trade-off value per the Eff-Tox Design	**-0.88**	**-0.57**	**-0.04**	**-0.24**	**-0.41**	**-**
	% selected by Eff-Tox Design	**0**	**0**	**41**	**46***	**11**	**2**
	% selected by OBD Isotonic Design	**3.5**	**2.1**	**39.4**	**47.2**	**7.8**	**0**
	% of times each dose level is acceptable for efficacy and toxicity in the ATLCEP Design (% of times each dose level is the one with the maximum value of the utility function when c = 1)	**23.03**** (**4.88**)***	**49.59** (**15.64**)	**95.51** (**58.81**)	**86.07** (**18.35**)	**29.61** (**2.13**)	**0.47**
2	True toxicity, true efficacy rate	0.20, 0.60	0.40, 0.62	0.55, 0.65	0.70, 0.70	0.85, 0.75	-
	Trade-off Value per the Eff-Tox Design	**-0.12**	**-0.40**	**-0.57**	**-0.70**	**-0.83**	**-**
	% selected by Eff-Tox Design	**74**	**16**	**2**	**1**	**0**	**7**
	% selected by OBD Isotonic Design	**64.9**	**32.3**	**2.6**	**0.2**	**0**	**0**
	% of times each dose level is acceptable for efficacy and toxicity in the ATLCEP Design (% of times each dose level is the one with the maximum value of the utility function when c = 1)	**93.77**** (**84.64**)***	**60.27** (**14.63**)	**5.59** (**0.71**)	**0.02** (**0.02**)	**0** (**0**)	**3.52**
3	True toxicity, true efficacy rate	0.30, 0.10	0.40, 0.30	0.55, 0.60	0.60, 0.62	0.65, 0.65	-
	Trade-off Value per the Eff-Tox Design	**-1.29**	**-1.04**	**-0.67**	**-0.71**	**-0.73**	**-**
	% selected by Eff-Tox Design	**0**	**3**	**4**	**0**	**0**	**93**
	% selected by OBD Isotonic Design	**49.1**	**39.6**	**9.1**	**1.8**	**0.3**	**0**
	% of times each dose level is acceptable for efficacy and toxicity in the ATLCEP Design	**9.30****	**13.04**	**4.08**	**0.15**	**0**	**77.75**

* The dose shown in underline is the dose level selected by each design as the optimal dose for toxicity and efficacy.

** These numbers for percentage of times each dose level is acceptable for toxicity and efficacy can add up to more than 100% in the ATLCEP design. For example, in scenario 1, dose levels 3 and 4 are selected in 96% and 86% of the simulation runs. This means that both dose levels 3 and 4 are selected in a large percentage of the 10000 simulations because both doses satisfy the Bayesian decision rules in those simulations.

*** The percentages shown in brackets are the percentages that each dose is chosen as the optimal dose for the ATLCEP design using the utility function and these percentages add up to 100.

Comparison of the ATLCEP Design to the Eff-Tox Design, the OBD Isotonic Design and to a 3+3 Phase I Design Followed by a Phase II Design

We compare, in Table 4, the performance of our proposed ATLCEP design to that of the Eff-Tox and the OBD Isotonic designs for various scenarios of true DLT and response rates. We use a sample size in the Eff-Tox design of 99 and cohort size of 9, which is the maximum cohort size allowed in the version of the Eff-Tox design software we used. We also use 0.33 for the upper limit of the probability of toxicity and 0.5 for the lower limit for the probability of efficacy, identical to the values we used in the decision rules for the ATLCEP design at the end of each simulated trial. All other input parameters used in the Eff-Tox and the OBD Isotonic design simulations are given in the Appendix.

For scenario 1 (Table 4), the Eff-Tox design does not select dose level 3, the dose with the highest trade-off value between the probabilities of efficacy and toxicity, as the optimal dose most frequently. The Eff-Tox design selects dose level 3 41% of the time, which is a little less frequent than it selects dose level 4, whose trade-off value is slightly lower. The OBD Isotonic design selects dose level 3, 39% of the time, while it selects dose level 4 47% of the time. The ATLCEP design selects dose level 3 and dose level 4 in 96% and 86% of the simulation runs respectively, as acceptable for toxicity and efficacy. From our simulations of the ATLCEP design, dose level 3 also has (a) the maximum value of the utility function most frequently (39%, 53% and 59% of the time for $c = 0.1$, $c = 0.5$ and $c = 1$ respectively), (b) the maximum value for the average percentage of patients with a response but no DLT and (c) the minimum value of the OR_i. Hence, based on these results of the ATLCEP design, our final selection for this design is dose level 3.

For scenario 2, the Eff-Tox design selects dose level 1, the dose with the highest trade-off value, 74% of the time and the OBD Isotonic design selects dose level 1 65% of the time as the optimal dose. The ATLCEP design selects dose level 1 as acceptable for safety and efficacy in 94% of the simulation runs. From our simulations of the ATLCEP design, dose level 1 also has (a) the maximum value of the utility function most frequently (53, 72 and 85% of the time for $c = 0.1$, $c = 0.5$ and $c = 1$ respectively), (b) the maximum value for the average percentage of patients with a response but no DLT and (c) the minimum value of the OR_i. Hence, based on these results of the ATLCEP design, our final selection for this design is dose level 1.

For scenario 3, the Eff-Tox design does not select any dose as optimal for safety and efficacy 93% of the time, while the ATLCEP design does not select any dose level as acceptable for safety and efficacy in 78% of the simulation runs. These results for the Eff-Tox design and

the ATLCEP design of selecting no dose level as optimal or acceptable for safety and efficacy most of the time are reasonable, since the true DLT rate at the lowest dose level itself is quite high - at a value of 0.3, it is just below the upper limit of 0.33 considered in the Bayesian decision rule for safety. The OBD Isotonic design does not perform well in this scenario since it assumes that at least the lowest dose is safe and selects dose level 1 as the optimal dose 49% of the time.

Although the comparisons between the ATLCEP design and the Eff-Tox and OBD Isotonic designs are not exact in terms of cohort size and sample size, these examples demonstrate that the ATLCEP design can select an optimal dose for efficacy and toxicity as robustly as the Eff-Tox and the OBD Isotonic designs.

We also compare the results of the ATLCEP design to those of a 3+3 Phase I design followed by a Phase II design. The 3+3 design picks the right dose for safety i.e., the true MTD of dose level 4 only 60% of the time for the true DLT rates in Table 1, as seen in Table 2. Hence the probability of selecting the right dose for both toxicity and efficacy at the end of Phase II, in a 3+3 followed by a Phase II design, is no more than 60%. In contrast, as seen in Table 3, the ATLCEP design picks dose level 4 as acceptable for safety and efficacy in 76% of the simulation runs (76% is also the value for the percentage of simulations that select dose level 4 as the optimal dose for toxicity and efficacy using the utility function for $c = 0.1$) for the true toxicity and true response rates in Table 1. Since the 3+3 design tends to stop earlier with a small sample size, its accuracy of MTD selection is very low. Our simulations with this and other scenarios confirm that we need a larger sample size to select the right dose with high accuracy before proceeding to the next study phase.

Discussion

Most Phase I dose-finding oncology trials enroll a very small number of patients and often fail to predict the MTD accurately due to the small sample size. We have shown, via simulations, that the accuracy of MTD or dose selection in these dose-finding designs increases considerably with an increase in sample size, and increases with an increase in cohort size in some cases and for some models. Thus, it is crucial to study the effect of sample size and cohort size on the accuracy of dose selection while designing an early phase oncology trial. With a larger number of patients, the efficacy of the drug can also be assessed in an early phase trial.

This has led us to propose a simple rule based design that enrolls a larger sample size than standard Phase 1 designs, which enables accurate dose selection with respect to toxicity and incorporates Bayesian decision rules for optimal dose selection for safety and efficacy at the end of the trial. We propose the *Accelerated Titration*

Large Cohort Early Phase (ATLCEP) design, a moderately sized, integrated Phase I/II trial design that assesses both safety and efficacy. We note that such a design should not become too large; the drawbacks of large seamless Phase I/II trials with more than a few hundred patients have been discussed by Mullard (2016). The ATLCEP design is intended to quickly move up the dose levels through accelerated titration but to have large enough sample sizes for doses near the MTD to substantially increase the accuracy of MTD selection and to provide assessment of efficacy in treatment response. The stopping rules and their timing used in our proposed ATLCEP design can be altered to create a modified ATLCEP design. For example, one could implement stopping if there are >= 3 DLTs, rather than >= 4 DLTs, in the first 6 patients in a dose level, or one could apply stopping rules after patient 7, 14, 20 etc. instead of after patient 6, 14, 20. However, additional simulations to study the operating characteristics of such a modified design would need to be performed. In this context, we note that there is a trade-off for using restrictive stopping rules and stopping too early for toxicity – it decreases the probability of identifying the optimal dose. We further note that a safety committee that can stop the trial at any point should be implemented with any of these designs. In summary, one needs to strike a balance between the unmodified 20+20 design, which enrolls 20 patients at once at a dose level and has a high accuracy of dose selection but exposes too many patients to the study drug at once and a design where few patients are enrolled at a time with stopping rules after each enrollment, for a maximum of 20-40 patients per dose level, which will have a lower accuracy of dose selection. Our proposed design falls in between these two extremes and provides a general framework for a larger A+B design that can be tweaked with stopping rules that are optimal for the study and study drug. This allows a trade-off to be struck between the safety of patients in the trial versus the accuracy of optimal dose selection for the safety and benefit of future patients.

The Bayesian decision rules used in the Eff-Tox design to optimize efficacy and safety in dose selection for each new cohort of patients are also used in the ALTCEP design, but only at the end of the trial. Our simulations comparing the performance of the ATLCEP design to that of the Eff-Tox and the OBD Isotonic design show that our design can perform as well as or better than these designs in the situations we considered. Also, this simple rule-based design can be implemented more easily than the Eff-Tox design and the OBD Isotonic design which require more advanced statistical calculations at each dosing stage.

We have also compared the results from the ATLCEP design to those from a 3+3 Phase I design followed by a Phase II design. Our simulations confirm that we need a larger sample size to select the right dose with high accuracy before proceeding to the next

study phase. The importance of selecting the optimal dose for toxicity and efficacy has previously been illustrated by Markman (2010) with the drug PLD for treatment of platinum resistant ovarian cancer. The dose of PLD in initial clinical trials which became the FDA approved dose proved to be too toxic and over time a lower dose has become the standard. We hope that the use of the ATLCEP design can help prevent such experiences in the future.

Conclusion

We propose the Accelerated Titration Large Cohort Early Phase (ATLCEP) design, a moderately large, simple rule-based integrated Phase I/II trial design that evaluates both safety and efficacy. This design incorporates stopping rules within dose levels to allow more flexible decision-making. We compare the operating characteristics of this design with other Phase I/II strategies, via simulations. Our simulations show that the design can perform as well as or better than the Eff-Tox or the Optimal Biological Dose (OBD) Isotonic design. It also performs better than a 3+3 Phase I design followed by a standard Phase II design.

Acknowledgment

Author's Contributions

RA contributed to the following: conception and design of the study, simulations and analysis, interpretation of the results, drafting of the article, and approval of the final submitted version, while SG, MC, GD, JM, DL and ML contributed to the conception and design of the study, helped with the interpretation of the results, provided critical reviews of drafts and approved the final submitted version.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the authors have read and approved the manuscript and that no ethical issues are involved.

References

Ananthakrishnan, R., S. Green, M. Chang, G. Doros and J. Massaro *et al*., 2017. Systematic comparison of the statistical operating characteristics of various Phase I oncology designs. Contemporary Clin. Trials Commun., 5: 34-48.
DOI: 10.1016/j.conctc.2016.11.006

Braun, T.M., 2002. The bivariate continual reassessment method: Extending the CRM to phase I trials of two competing outcomes. Control Clin. Trials, 23: 240-256. DOI: 10.1016/S0197-2456(01)00205-7

Cai, C., Y. Yuan and Y. Ji, 2014. A Bayesian dose finding design for oncology clinical trials of combinational biological agents. J. R Stat. Society Ser., 1: 159-173. DOI: 10.1111/rssc.12039

Dragalin, V., V.V. Fedorov and Y. Wu, 2008. Two-stage design for dose-finding that accounts for both efficacy and safety. Stat. Med., 10: 5156-5176. DOI: 10.1002/sim.3356

Ferrandina, G., 2010. Reply to M. Markman. J. Clin. Oncol., 1: e321-e322. DOI: 10.1200/JCO.2010.29.0866

Garrett-Mayer, E., 2006. The continual reassessment method for dose-finding studies: A tutorial. Clin. Trials, 3: 57-71. DOI: 10.1191/1740774506cn134oa

Gezmu, M. and N. Flournoy, 2006. Group up-and-down designs for dose-finding. J. Stat. Plann. Inference, 136: 1749-1764. DOI: 10.1016/j.jspi.2005.08.002

Hansen, A.R., D.M. Graham, G.R. Pond and L.L. Siu, 2014. Phase 1 trial design: Is 3 + 3 the best? Cancer Control, 21: 200-208. DOI: 10.1177/107327481402100304

Hoering, A., A. Mitchell, M. LeBlanc and J. Crowley, 2013. Early phase trial design for assessing several dose levels for toxicity and efficacy for targeted agents. Clin. Trials, 10: 422-429. DOI: 10.1177/1740774513480961

Hoering, A., M. LeBlanc and J. Crowley, 2011. Seamless phase I-II trial design for assessing toxicity and efficacy for targeted agents. Clin. Cancer Res., 15: 640-646. DOI: 10.1158/1078-0432.CCR-10-1262

Huang, B., P. Bycott and E. Talukder, 2016. Novel dose-finding designs and considerations on practical implementations in oncology clinical trials. J. Biopharm. Stat., 16: 1-12. DOI: 10.1080/10543406.2016.1148715

Ivanova, A., 2006a. Escalation, group and A + B designs for dose-finding trials. Stat. Med., 25: 3668-3678. DOI: 10.1002/sim.2470

Ivanova, A., 2006b. Dose-Finding in Oncology: Nonparametric Methods. In: Dose Finding in Drug Development, Ting, N. (Ed.), Springer, New York, pp: 49-58

Ivanova, A., K. Liu, E. Snyder and D. Snavely, 2009. An adaptive design for identifying the dose with the best efficacy/tolerability profile with application to a crossover dose-finding study. Stat. Med., 28: 2941-2951. DOI: 10.1002/sim.3684

Li, D.H., J.B. Whitmore, W. Guo and Y. Ji, 2017. Toxicity and efficacy probability interval design for Phase I adoptive cell therapy dose-finding clinical trials. Clin. Cancer Res., 23: 13-20. DOI: 10.1158/1078-0432.CCR-16-1125

Markman, M., 2010. Serious ethical dilemma of single-agent pegylated liposomal doxorubicin employed as a control arm in ovarian cancer chemotherapy trials. J. Clin. Oncol., 28: e319-e320. DOI: 10.1200/JCO.2010.28.9934

Mullard, A., 2016. Reining in the supersized Phase I cancer trial. Nat. Rev. Drug Discov., 15: 371-373. DOI: 10.1038/nrd.2016.110

O'Quigley, J., M. Pepe and L. Fisher, 1990. Continual reassessment method: A practical design for phase 1 clinical trials in cancer. Biometrics, 46: 33-48. DOI: 10.2307/2531628

Pan, H., F. Xie, P. Liu, J. Xia and Y. Ji, 2014. A phase I/II Seamless dose Escalation/expansion with Adaptive Randomization Scheme (SEARS). Clin. Trials, 11: 49-59. DOI: 10.1177/1740774513500081

Pan, H., P. Huang, Z. Wang, L. Wang and C. Li *et al.*, 2013. A novel Bayesian seamless phase I/II design. PLoS One, 8: e73060-e73060. DOI: 10.1371/journal.pone.0073060

Sato, H., A. Hirakawa and C. Hamada, 2016. An adaptive dose-finding method using a change-point model for molecularly targeted agents in phase I trials. Stat. Med., 35: 4093-4109. DOI: 10.1002/sim.6981

Schilsky, R.L., L. Minasian, D. Auclair, A. Rahman and R. Pazdur, 2013. Optimizing dosing of oncology drugs. Proceedings of the Conference on Clinical Cancer Research, (CCR' 13).

Thall, P.F. and J.D. Cook, 2004. Dose-finding based on efficacy-toxicity trade-offs. Biometrics, 60: 684-693. DOI: 10.1111/j.0006-341X.2004.00218.x

Thall, P.F., J.D. Cook and E. Estey, 2006. Adaptive dose selection using efficacy-toxicity trade-offs: Illustrations and practical considerations. J. Biopharmaceutical Stat., 16: 623-638. DOI: 10.1080/10543400600860394

Wages, N.A. and C. Tait, 2015. Seamless Phase I/II adaptive design for oncology trials of molecularly targeted agents. J. Biopharm. Stat., 25: 903-920. DOI: 10.1080/10543406.2014.920873

Yang, S., S. Wang and Y. Ji, 2015. An integrated dose-finding tool for phase I trials in oncology. Contemporary Clin. Trials, 45: 426-434. DOI: 10.1016/j.cct.2015.09.019

Yin, G., Y. Li and Y. Ji, 2006. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. Biometrics, 62: 777-784. DOI: 10.1111/j.1541-0420.2006.00534.x

Yuan, Y., H.Q. Nguyen and P.F. Thall, 2016. Bayesian Designs for Phase I-II Clinical Trials. 1st Edn., Chapman and Hall/CRC, ISBN-10: 1498709559, pp: 310.

Zang, Y., J.J. Lee and Y. Yuan, 2014. Adaptive designs for identifying optimal biological dose for molecularly targeted agents. Clin. Trials, 11: 319-327. DOI: 10.1177/1740774514529848

Zhang, W., D.J. Sargent and S. Mandrekar, 2006. An adaptive dose-finding design incorporating both toxicity and efficacy. Stat. Med., 25: 2365-2383. DOI: 10.1002/sim.2325

Appendix:

Target DLT interval of the 20+20 design

We have constructed the stopping rules for the 20+20 design such that it targets a DLT rate Γ of ~0.2, similar to the 3+3 design, which targets an approximate DLT rate of 0.2 (range of 0.17-0.26) (Ivanova, 2006a; Gezmu and Flournoy, 2006). Other stopping rules and cohort sizes can be proposed in order to target other DLT rates. The approximate DLT interval that any A+B design targets can be calculated using the following inequality from Ivanova (2006a):

$$C_U / (A + B) < \Gamma < \Gamma_{A+B}$$

Where:

Γ_{A+B} = The solution to the equation $\Pr(\text{Bin}(A+B, \Gamma_{A+B}) \leq C_U) = 0.5$

$A+B$ = The total number of patients a dose level can enroll and the trial is stopped if $>C_U$ DLTs are observed in $A+B$ patients

For our 20+20 design, $A+B = 40$, $C_U = 8$ and the inequality becomes:

$$0.2 < \Gamma < 0.21$$

Correlation Coefficient r that Can Be Used (for Different Cases of True Efficacy and Toxicity Rates)

Dose Level	True toxicity rate	True efficacy rate
1	0.01	0.05
2	0.02	0.25
3	0.06	0.30
4	0.20	0.35
5	0.55	0.40
>= 6	0.89	0.50

Maximum positive value of r in this case where both the true DLT and response rates are increasing monotonically with an increase in dose is ~0.25.

Dose Level	True toxicity rate	True efficacy rate
1	0.01	0.4
2	0.02	0.35
3	0.06	0.30
4	0.20	0.25
5	0.55	0.15
>=6	0.89	0.05

Maximum positive value of *r* in this case where the true DLT rate is increasing monotonically with an increase in dose but the true response rate is monotonically deceasing is 0.08.

Input Parameters Used in Simulations for the CRM, Eff-Tox and OBD Isotonic Designs

Input Parameters Used in the R Package CRM for the CRM Design

A CRM design with a target DLT rate of 0.2, starting dose level of 1 and a 1-parameter logistic dose-toxicity model with parameter "a" whose initial value is 1 and fixed parameter "b" whose value is 3 is considered. The prior for "a" is exp(-a). The prior DLT rate at each of the six dose levels is (0.15, 0.25, 0.3, 0.45, 0.51, 0.56) and the true DLT rate at each dose level is as given in Table 1.

Input Parameters Used in the Eff-Tox Package for the Eff-Tox Design

Starting dose = 1
Cohort size = 9
Number of cohorts = 11
Number of simulations=10000

Probability of Toxicity and Efficacy Limits for Dose Acceptability Rules

Parameter	Value
Prob(tox) upper limit (π_T*)	0.33000
Lower prob cutoff for prob of toxicity ($p_{T,L}$)	0.10000
Prob(eff) lower limit (π_E*)	0.50000
Lower prob cutoff for prob of efficacy ($p_{E,L}$)	0.10000

Trade-off Function Elicited Points (3 points to define the trade-off function contour)

	π_E	π_T
($\pi_{1,E}$*, 0)	0.50000	0.00000
(1, $\pi_{2,T}$*)	1.00000	0.65000
($\pi_{3,E}$, $\pi_{3,T}$)	0.70000	0.25000

Elicited Means (Prior Toxicity, Prior Efficacy)

Dose	Toxicity	Efficacy
1	0.0200	0.2000
2	0.0400	0.4000
3	0.0600	0.6000
4	0.0800	0.8000
5	0.1000	0.9000

Number of simulations = 10000
Input parameters Used in the OBD Isotonic Design
Starting dose =1
Cohort size = 9
Number of cohorts = 11
phi = upper bound of toxicity rate = 0.33
ct = threshold for posterior probability of toxicity (any dose with toxicity probability larger than ct is excluded from the admissible set of doses) = 0.9
Number of simulations = 10000

The R code given at the following URL was used along with the input parameters given above: http://odin.mdacc.tmc.edu/~yyuan/Software/TargetAgent/targetAgentDF.r
The parameters in the beta prior distribution for toxicity response probability were changed in the R code above to reflect the values of ct (= 0.9) and phi (= 0.33) used in our simulations.

Results for the ATLCEP Design for Various Scenarios of True Toxicity and Efficacy Rates (Correlation Coefficient r=0 for all the below simulations)

(a) True toxicity and efficacy rates are monotonically increasing with an increase in dose

	Dose Level 1	Dose Level 2	Dose Level 3	**Dose Level 4**	Dose Level 5	Dose Level 6
True toxicity, true efficacy rate	0.01, 0.01	0.02, 0.05	0.06, 0.15	0.2, 0.45	0.55, 0.5	0.89, 0.6
Odds of Toxicity to Efficacy from Simulations at Dose Level i (OR_i)*	0.20	0.26	0.35	0.31	2.6	3.7
Average Number of Patients	3.5	4.5	7.3	14.0	12.2	0.28
Average Number of DLTs	0.04	0.1	0.45	2.8	6.7	0.25
Average Number of Responses	0.04	0.22	1.1	6.3	6.1	0.17
Average Number of Patients with a Response	0.04	0.22	1.0	5.0	2.7	0.02

but no DLT						
Decision based on the Bayesian Posterior Probabilities (% of times out of 10000 simulations each dose level satisfies the Bayesian decision criteria for both toxicity and efficacy)**	2.86%	13.02%	28. 6%	76.08%	15.32%	0%
% of times out of 10000 simulations each dose level is the one with the maximum value of the utility function						
when c = 0.1***	0.31%	1.61%	6.43%	47.01%	42.91%	1.73%
when c = 0.5	2.13%	5.64%	14.04%	62.19%	15.63%	0.37%
when c = 1	6.05%	8.48%	20.21%	61.13%	4.11%	0.02%

*Calculated using only those simulations runs with a non-zero denominator for OR_i. For the lower dose levels (levels 1 and 2), the denominator is zero in many simulation runs since the average number of responses is zero.

**No dose level is chosen as acceptable for toxicity and efficacy~13% of the time. The addition of the percentages for dose selection based on the Bayesian decision rules can add up to more than 100 since more than one dose level can be chosen as acceptable for toxicity and efficacy in each simulation.

***c = 1 gives equal weight to toxicity and efficacy while 0.1 gives a very small weight to toxicity and more weight to efficacy.

Mean sample size for this example is 41.75; median is 35; minimum is 12 and maximum is 126.

Dose level selected as optimal by this design is shown in bold.

(b) True Toxicity Rates are Monotonically Increasing with an Increase in Dose and True Efficacy Rates Increase with an Increase in Dose but Plateau after Dose Level 3

	Dose Level 1	Dose Level 2	Dose Level 3	**Dose Level 4**	Dose Level 5	Dose Level 6
True toxicity, true efficacy rate	0.01, 0.01	0.02, 0.05	0.06, 0.15	0.2, 0.45	0.55, 0.45	0.89, 0.45
Odds of Toxicity to Efficacy from Simulations at Dose Level i (OR_i)*	0.20	0.26	0.35	0.31	3.13	6.95
Average Number of Patients	3.5	4.5	7.3	14.0	12.2	0.28
Average Number of DLTs	0.04	0.1	0.45	2.8	6.7	0.25
Average Number of Responses	0.04	0.22	1.1	6.3	5.5	0.12
Average Number of Patients with a Response but no DLT	0.04	0.22	1.0	5.0	2.4	0.01
Decision based on the Bayesian Posterior Probabilities (% of times out of 10000 simulations each dose level satisfies the Bayesian decision criteria for both toxicity and efficacy)**	2.86%	13.02%	28.6%	76.08%	13.57%	0.0%
% of times out of 10000 simulations each dose level is the one with the maximum value of the utility function						
when c =0.1***	0.46%	2.15%	7.32%	54.07%	35.03%	0.97%
when c = 0.5	2.45%	6.03%	14.71%	65.77%	10.94%	0.1%
when c = 1	6.43%	8.55%	20.44%	61.93%	2.65%	0%

* Calculated using only those simulations runs with a non-zero denominator for OR_i. For the lower dose levels (levels 1 and 2), the denominator is zero in many simulation runs since the average number of responses is zero.

**No dose level is chosen as acceptable for toxicity and efficacy ~13% of the time. The addition of the percentages for dose selection based on the Bayesian decision rules can add up to more than 100 since more than one dose level can be chosen as acceptable for toxicity and efficacy in each simulation.

***c = 1 gives equal weight to toxicity and efficacy while 0.1 gives a very small weight to toxicity and more weight to efficacy.

Mean sample size for this example is 41.75; median is 35; minimum is 12 and maximum is 126.

Dose level selected as optimal by this design is shown in bold.

(c)

	Dose Level 1	Dose Level 2	**Dose Level 3**	Dose Level 4	Dose Level 5	Dose Level 6
True toxicity, true efficacy rate	0.05, 0.1	0.15, 0.3	0.26, 0.45	0.38, 0.55	0.55, 0.62	0.7, 0.7
Odds of Toxicity to Efficacy from Simulations at Dose Level i (OR_i)*	0.42	0.4	0.56	0.87	1.48	1.83
Average Number of Patients	5.4	11.0	16.7	14.3	3.5	0.08
Average Number of DLTs	0.27	1.6	4.3	5.4	1.9	0.06
Average Number of Responses	0.54	3.3	7.6	7.9	2.2	0.05
Average Number of Patients with a Response but no DLT	0.51	2.8	5.6	4.9	0.98	0.02
Decision based on the Bayesian Posterior Probabilities (% of times out of 10000 simulations each dose level satisfies the Bayesian decision criteria for both toxicity	23.21%	44.51%	72.03%	51.97%	4.56%	0.04%

and efficacy)**
% of times out of 10000 simulations each
dose level is the one with the maximum
value of the utility function

when c = 0.1***	3.42%	15.40%	34.51%	35.25%	11.14%	0.28%
when c = 0.5	9.24%	23.71%	36.84%	25.65%	4.49%	0.07%
when c = 1	18.21%	32.24%	34.63%	14.01%	0.9%	0.01%

*Calculated using only those simulations runs with a non-zero denominator for OR_i. For the lower dose levels (levels 1 and 2), the denominator is zero in many simulation runs since the average number of responses is zero.

No dose level is chosen as acceptable for toxicity and efficacy ~8%** of the time. The addition of the percentages for dose selection based on the Bayesian decision rules can add up to more than 100 since more than one dose level can be chosen as acceptable for toxicity and efficacy in each simulation.

***c = 1 gives equal weight to toxicity and efficacy while 0.1 gives a very small weight to toxicity and more weight to efficacy.

Mean sample size for this example is 51.12; median is 49; minimum s 6 and maximum is 126.

Dose level selected as optimal by this design is shown in bold.

(d)

Scenario		Dose Level					None
		1	2	3	4	5	
1	True toxicity, true efficacy rate	0.05, 0.10	0.10, 0.25	**0.3, 0.48**	0.45, 0.55	0.6, 0.3	-
	% of times each dose level is acceptable for efficacy and toxicity in the ATLCEP Design (% of times each dose level is the one with the maximum value of the utility functionwhen c = 1)	23 (20.16)	40.09 (33.94)	75.37 (39)	30.71 (6.88)	0.38 (0.02)	9.5
2	True toxicity, true efficacy rate	0.05,0.1	**0.25, 0.45**	0.4, 0.56	0.55, 0.3	0.6, 0.1	-
	% of times each dose level is acceptable for efficacy and toxicity in the ATLCEP Design (% of times each dose level is the one with the maximum value of the utility function when c =1)	21.84 (22.94)	74.61 (57.13)	51.74 (19.9)	1.37 (0.03)	0.03 (0.0)	11.09
3	True toxicity, true efficacy rate	0.01,0.1	0.02, 0.35	**0.06 0.5**	0.2, 0.3	0.55, 0.2	-
	% of times each dose level is acceptable for efficacy and toxicity in the ATLCEP Design (% of times each dose level is the one with the maximum value of the utility function when c = 1)	27.09 (9.91)	70.14 (37.05)	86.85 (47.76)	38.55 (5.08)	1.95 (0.18)	1.83

Dose level selected as optimal by this design is shown in bold in the first row of each scenario