

Toxicity and Efficacy Probability Interval Design for Phase I Adoptive Cell Therapy Dose-Finding Clinical Trials

Daniel H. Li¹, James B. Whitmore², Wentian Guo³, and Yuan Ji⁴

Abstract

Recent trials of adoptive cell therapy (ACT), such as the chimeric antigen receptor (CAR) T-cell therapy, have demonstrated promising therapeutic effects for cancer patients. A main issue in the product development is to determine the appropriate dose of ACT. Traditional phase I trial designs for cytotoxic agents explicitly assume that toxicity increases monotonically with dose levels and implicitly assume the same for efficacy to justify dose escalation. ACT usually induces rapid responses, and the monotonic dose–response assumption is unlikely to hold due to its immunobiologic activities. We propose a toxicity and efficacy probability interval (TEPI) design for dose finding in ACT trials. This approach incorpo-

rates efficacy outcomes to inform dosing decisions to optimize efficacy and safety simultaneously. Rather than finding the maximum tolerated dose (MTD), the TEPI design is aimed at finding the dose with the most desirable outcome for safety and efficacy. The key features of TEPI are its simplicity, flexibility, and transparency, because all decision rules can be prespecified prior to trial initiation. We conduct simulation studies to investigate the operating characteristics of the TEPI design and compare it to existing methods. In summary, the TEPI design is a novel method for ACT dose finding, which possesses superior performance and is easy to use, simple, and transparent. *Clin Cancer Res*; 23(1); 13–20. ©2016 AACR.

Introduction

In the past few years, promising antitumor effects have been seen in patients with late-stage cancer when they were treated with adoptive cell therapy (ACT), including chimeric antigen receptor (CAR) T cells, T-cell receptors (TCR), and tumor-infiltrating lymphocytes (TIL). Early data continue to warrant the expedited development of these therapies as a treatment option for patients with various cancers. ACT is a highly personalized cancer therapy that harnesses a patient's own immune cells (specifically, T lymphocytes) to recognize cancer-specific abnormalities, enabling them to target and attack malignant cells throughout the body. However, challenging questions remain regarding the immunobiology and development of these personalized therapies, such as the design of an appropriate phase I dose-escalation study that takes into consideration the unique properties of ACTs.

Traditional phase I dose-finding oncology trials aim to identify the maximum tolerated dose (MTD), which is the highest dose that has a dose-limiting toxicity (DLT) rate less than or close to a

prespecified target rate, say 30%. The commonly used methods are the rule-based designs, such as the 3 + 3 design (1), and the model-based designs, such as the modified toxicity probability interval (mTPI) design (2–4) and the continual reassessment method (CRM; ref. 5).

Traditional designs considering the DLT data implicitly assume a monotonically increasing relationship between dose and response efficacy; otherwise, there is no justification to escalate the dose level when it is safe. Monotone efficacy may be a reasonable assumption for cytotoxic agents. However, this assumption may not hold for ACTs. In ACT trials, the MTD is not always optimal, as clinical response correlates less with dose level. For example, recent phase I trials of CAR T cells targeting CD19 suggest that a range of dose levels is generally safe and effective, and there is no clear correlation between T-cell dose and clinical response (6). In particular, two patients in the CAR T-cell trials reported in refs. 7 and 8 exhibited inferior efficacy outcomes to another patient despite receiving a 10-fold higher T-cell dose. A TIL study for metastatic cancer (9) found no correlation between the number of cells administered and the likelihood of a clinical response. Therefore, traditional algorithmic and model-based designs based on binary measures of toxicity may be inappropriate for identifying the optimal dose of ACTs. Given the complexity of an ACT product, preliminary dose exploration should aim to capture effective biologic activity rather than dose-limiting toxicity alone (10).

Emerging data suggest that CAR T-cell therapy in B-cell hematologic malignancies may induce rapid responses. Toxicity and efficacy of a biomarker (e.g., cell expansion) may, therefore, be measured in the same time frame (6, 11, 12). Model-based methods have been developed to model toxicity and efficacy data jointly in order to determine the acceptable dose level (13); these

¹Juno Therapeutics, Seattle, Washington. ²Juno Therapeutics, Seattle, Washington. ³Laiya Consulting, Inc., Wilmette, Illinois. ⁴NorthShore University HealthSystem, Chicago, Illinois.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Yuan Ji, NorthShore University HealthSystem/The University of Chicago, 1001 University Place, Evanston, IL 60201. Phone: 224-364-7312; Fax: 224-364-7319; E-mail: koaeraser@gmail.com

doi: 10.1158/1078-0432.CCR-16-1125

©2016 American Association for Cancer Research.

methods are powerful and effective, although they may require substantial investigation from trial statisticians to ensure proper implementation and calibration. To this end, we developed a practical dose-finding design for an ACT that (i) incorporates both toxicity and efficacy data; (ii) provides the same adaptive feature as model-based designs; and (iii) most importantly, is transparent to clinicians and as simple to implement as the 3 + 3 and mTPI designs. We propose a toxicity and efficacy probability interval (TEPI) design, which is based on a clinician-elicited decision table in terms of efficacy and toxicity probability intervals. This design is motivated for the conduct of a phase I CD19-targeted CAR T-cell therapy dose-finding trial.

Materials and Methods

Elicited decision table

Consider d ascending doses in a single-agent ACT phase I trial. Due to ethical considerations, it is always assumed that the toxicity probability p_i increases with dose level i . However, the efficacy probability q_i may increase initially and then reach a plateau from which minimal improvement or even decreasing efficacy may be seen with increasing dose. For this reason, we assume that q_i is not monotone with i , and that p_i and q_i are independent. Suppose that dose i is currently used in the trial and n_i patients have already been allocated to dose i , with x_i and y_i patients experiencing toxicity and efficacy outcomes. Aggregating across all the doses, the trial data are denoted as $D = \{(n_i, x_i, y_i), i = 1, \dots, d\}$.

Similar to mTPI, we partition the unit intervals $(0, 1)$ for p_i and q_i into subintervals. Denoting (a, b) and (c, d) a subinterval in the partition for p_i and q_i , respectively. The interval combination $(a, b) \times (c, d)$ forms the basis for dose-finding decisions, with each combination corresponding to a specific decision, such as dose escalation or de-escalation. A dose-finding decision table can then be elicited with clinicians for all interval combinations. An example of such a two-dimensional table is given in Table 1. We call this the "preset table" for the TEPI design, which is fixed and elicited prior to the trial. As illustrated in Table 1, there are four subintervals for toxicity (rows) and efficacy (columns), the intersection of which forms 16 interval combinations. Each of the 16 combinations corresponds to a specific dosing decision. Decision "E" denotes escalation (i.e., treating the next cohort of patients at the next higher dose). Decision "S" denotes staying at the current dose for treating the next cohort of patients. Decision "D" denotes de-escalation (i.e., treating the next cohort of patients at the next lower dose). These reflect practical clinical actions when the particular combination of toxicity and efficacy data are observed at a certain dose level. For example, Table 1 shows that the interval combination $(0, 0.15) \times (0, 0.2)$ for p_i and q_i corresponds to an action of "E"—escalation. This means that if the observed toxicity rate for a dose falls in $(0, 0.15)$ and the observed efficacy rate is in $(0, 0.2)$, the next patient cohort will

be treated at the next higher dose level. In order to formulate this table, it is required to have determined: (i) the maximum tolerated toxicity rate, p_T , and (ii) the minimum acceptable efficacy rate, q_E , at which the clinician is willing to treat future patients at the current dose level.

Table 1 is elicited for an ACT trial based on $p_T = 0.4$, $q_E = 0.2$, historical data for the ACT, and clinician input. For other trials, the preset table can be modified accordingly. A more detailed description of Table 1 is given in Supplementary Material A.

Dose-finding algorithm

Building upon the preset table, we set up a "local" decision-theoretic framework and derive a Bayes rule. Here, local means that the framework focuses on the optimal decision to be made for the current dose instead of the trial. We show that the Bayes rule is equivalent to computing the joint unit probability mass (JUPM) for the toxicity and efficacy probability intervals. For a given region A , the JUPM is defined as the ratio between the probability of the region and the size of the region (2, 14). Considering the two-dimensional unit square $(0, 1) \times (0, 1)$ in the real space, the JUPM for each interval combination $(a, b) \times (c, d)$ is

$$JUPM_{(a,b),(c,d)}^{(c,d)} \equiv \frac{Pr\{p_i \in (a, b), q_i \in (c, d) | D\}}{(b-a) \times (d-c)}, \quad 0 < a < b < 1, 0 < c < d < 1$$

Here, the numerator, $Pr\{p_i \in (a, b), q_i \in (c, d) | D\}$, is the posterior probability of p_i and q_i falling in the interval (a, b) and (c, d) , respectively.

Assume the prior for each p_i follows an independent $beta(\alpha_p, \beta_p)$, and the prior for each q_i follows an independent $beta(\alpha_q, \beta_q)$, where $beta(\alpha, \beta)$ denotes a beta distribution with mean $\alpha/(\alpha + \beta)$. The rationale of using independent priors follows the same argument in (2). Assume $x_i | p_i : Bin(n_i, p_i)$ and $y_i | q_i : Bin(n_i, q_i)$, where $Bin(n, q)$ denotes a binomial distribution with n trials and q probability of success. Then, the likelihood function for the observed toxicity data $(x_i, n_i), i = 1, \dots, d$, is a product of binomial densities $l(p) = \prod_{i=1}^d p_i^{x_i} (1 - p_i)^{n_i - x_i}$, and the likelihood function for the efficacy data $(y_i, n_i), i = 1, \dots, d$ is a product of binomial densities $l(q) = \prod_{i=1}^d q_i^{y_i} (1 - q_i)^{n_i - y_i}$. Thus, the posterior distributions for p_i and q_i are $beta(\alpha_p + x_i, \beta_p + n_i - x_i)$ and $beta(\alpha_q + y_i, \beta_q + n_i - y_i)$, respectively. Based on the posterior distributions, there exists a "winning" interval combination $(a^*, b^*) \times (c^*, d^*)$ that achieves the maximum JUPM among all the combinations in Table 1, and the corresponding decision for that combination is selected for treating the next cohort of patients. It can be shown that the decision is the Bayes rule under a balanced loss function (Supplementary Material B).

The basic dose-finding concept of TEPI is as follows. Assume that the current patient cohort is treated at dose i . After the current cohort completes DLT and response evaluation, compute the

Table 1. An example of a probability decision table based on $p_T = 0.4$ and $q_E = 0.2$

			Efficacy rate			
			Low (0, 0.2)	Moderate (0.2, 0.4)	High (0.4, 0.6)	Superb (0.6, 1)
Toxicity rate	Low	(0, 0.15)	E	E	E	E
	Moderate	(0.15, 0.33)	E	E	E	S
	High	(0.33, 0.4)	D	S	S	S
	Unacceptable	(0.4, 0.1)	D	D	D	D

NOTE: "E," "S," and "D" denote escalation, stay, and de-escalation, respectively.

Table 2. Dose-finding table of the TEPI design

Number of patients treated at current dose level	Number of DLTs	Number of responders			
		0	1-3	5-6	7-9
3	0	E	E		
	1	D	S		
	2	D	D		
	3	DU _T	DU _T		
		0	1-4	5-6	
6	0	EU	E	E	
	1	EU	E	S	
	2-3	DU _E	D	S	
	4	DU _E	D	D	
	5-6	DU _T	DU _T	DU _T	
	0	1	2-6	7-9	
9	0-1	EU	E	E	E
	2	EU	E	E	S
	3-4	DU _E	D	S	S
	5-6	DU _E	D	D	D
	7-9	DU _T	DU _T	DU _T	DU _T
	0-1	2	3-7	8-12	
12	0-1	EU	E	E	E
	2	EU	E	E	S
	3-5	DU _E	D	S	S
	6	DU _E	D	D	D
	7-12	DU _T	DU _T	DU _T	DU _T
	0-1	2	3-9	10-15	
15	0-2	EU	E	E	E
	3-4	EU	E	E	S
	5-7	DU _E	D	S	S
	8-9	DU _E	D	D	D
	10-15	DU _T	DU _T	DU _T	DU _T
	0-2	3	4-11	12-18	
18	0-2	EU	E	E	E
	3-5	EU	E	E	S
	6-9	DU _E	D	S	S
	10	DU _E	D	D	D
	11-18	DU _T	DU _T	DU _T	DU _T
	0-2	3	4-13	14-21	
21	0-2	EU	E	E	E
	3-6	EU	E	E	S
	7-10	DU _E	D	S	S
	11-12	DU _E	D	D	D
	13-21	DU _T	DU _T	DU _T	DU _T
	0-3	4	5-15	16-24	
24	0-3	EU	E	E	E
	4-6	EU	E	E	S
	7-12	DU _E	D	S	S
	13	DU _E	D	D	D
	16-24	DU _T	DU _T	DU _T	DU _T
	0-3	4-5	6-17	18-27	
27	0-3	EU	E	E	E
	4-7	EU	E	E	S
	8-13	DU _E	D	S	S
	14	DU _E	D	D	D
	16-24	DU _T	DU _T	DU _T	DU _T

NOTE: The letters are computed based on the preset decisions in Table 1 and represent different dose-finding actions during the trial based on the observed toxicity and efficacy data. Decisions "D," "S," and "E" correspond to actions of de-escalate, stay (retain), and escalate the dose level, respectively. Decisions "DU_T" and "DU_E" correspond to actions of de-escalate the dose level due to high toxicity or low efficacy, respectively, and to mark the current dose unacceptable for future use. Decision "EU" is to escalate the dose level and mark the current dose unacceptable for future use.

JUPMs for all the interval combinations in Table 1. The TEPI design recommends "E," "S," or "D" corresponding to the combination with the largest JUPM value. Because for a given trial there are a finite number of possible toxicity and efficacy out-

comes as binomial counts, for any toxicity and efficacy counts that can be observed in the trial, the TEPI dose-finding decisions can be precalculated. For our ACT trial, based on Table 1, all the decisions have been precalculated and presented in Table 2. Similar to the

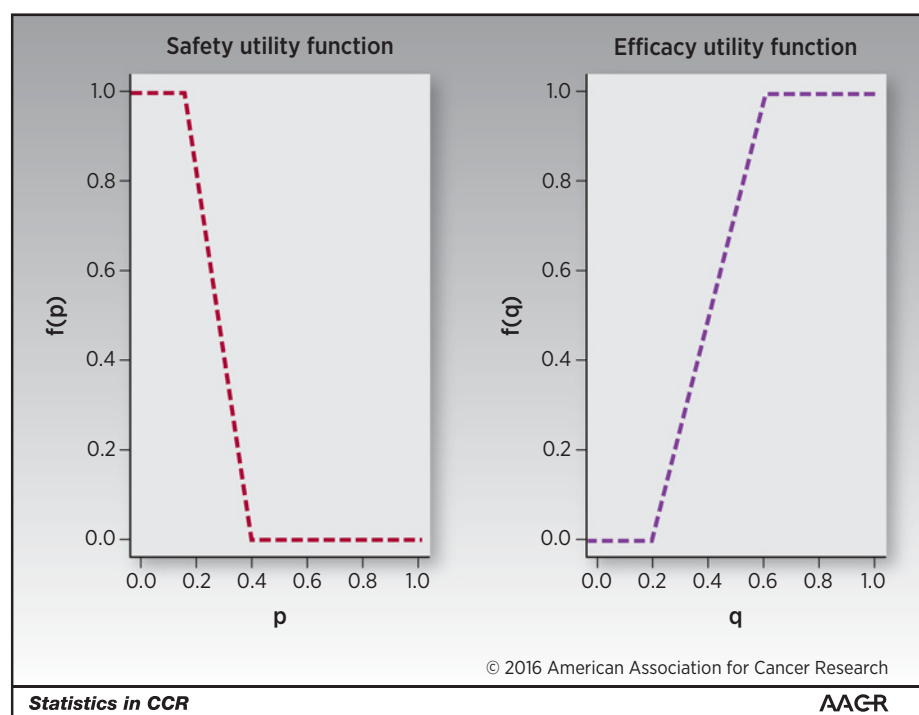


Figure 1.
Safety and efficacy utility functions.

decision table for mTPI, this table enables clinicians to conduct the trial with transparency.

In practice, the TEPI design needs to be calibrated according to physicians' needs. This is transparent and requires little effort. The tuning is for the intervals in Table 1 so that the decisions in Table 2 are satisfactory to the clinicians. Specifically, by modifying the interval combinations in Table 1, a new decision table can be generated in the form of Table 2. The tuning is completed once the desirable decisions are obtained.

To enable ethical constraints, we introduce two additional rules as part of the dose-finding algorithm. One is to exclude any dose with excessive toxicity, and the other is to exclude any dose with unacceptable efficacy.

- **Safety rule:** If $\Pr(p_i > p_T | D) > \eta$ for a η close to 1 (say, 0.95), exclude dose $i, i + 1, \dots, d$ from future use for this trial (i.e., these doses will never be tested again in the trial) and treat the next cohort of patients at dose $i - 1$. This corresponds to a dosing action of "DU_T"—de-escalate due to unacceptable high toxicity.
- **Futility rule:** If $\Pr(q_i > q_E | D) < \xi$ for a small ξ (say 0.3), then exclude dose i from future use in the trial. This corresponds to a dosing action of "EU"—escalate and never return due to unacceptable low efficacy—or "DU_E"—de-escalate and never return this dose due to unacceptable low efficacy.

A dose level is considered "available" if it satisfies both the safety and futility rules, and only these doses can be used to treat subsequent patients.

Final dose selection

At the end of the trial, we select the most desirable dose based on a utility score to balance the toxicity and efficacy tradeoff. Utility-based decision criteria have been adopted widely in recent dose-finding trials (15–17). An elicited utility function for safety

and efficacy can be constructed based on p_T and q_E through discussions with clinicians. For example, Fig. 1 (left) shows the utility function $f_1(p)$ for safety, where p denotes the toxicity rate. Utility $f_1(p)$ is set to 1 if $p \leq 15\%$, 0 if $p > 40\%$, and linearly decreasing with p if $p \in (15\%, 40\%)$. Utility $f_2(q)$ is set in a similar fashion. Figure 1 shows both utility functions. We assume a monotonic constraint on priors for p_i 's while selecting the best dose (i.e., $p_1 \leq p_2 \leq \dots \leq p_d$). The utility score function is defined as $U(p, q) = f_1(p)f_2(q)$, where p denotes the toxicity rate, and q denotes the efficacy rate. Both $f_1(\cdot)$ and $f_2(\cdot)$ are truncated linear functions, given by

$$f_1(p) = \begin{cases} 1, & p \in (0, p_1^*], \\ 1 - \frac{p - p_1^*}{p_2^* - p_1^*}, & p \in (p_1^*, p_2^*), \\ 0, & p \in [p_2^*, 1), \end{cases} \quad f_2(q) = \begin{cases} 0, & q \in (0, q_1^*], \\ \frac{q - q_1^*}{q_2^* - q_1^*}, & q \in (q_1^*, q_2^*), \\ 1, & q \in [q_2^*, 1), \end{cases}$$

where p^* 's and q^* 's are prespecified cutoff values. For each dose i , we use a simple numerical approximation approach to compute the posterior expected utility, $E[U(p_i, q_i) | D]$. We generate a total of T random samples from the posterior distributions. For each sample t , we generate $p^t = (p_1^t, \dots, p_d^t)$ and $q^t = (q_1^t, \dots, q_d^t)$ as a random sample of d probabilities of toxicity and efficacy, respectively. We perform the isotonic transformation (2, 14) on p^t to obtain $\hat{p}^t = (\hat{p}_1^t, \dots, \hat{p}_d^t)$ where $\hat{p}_i^t \leq \hat{p}_j^t$ if $i < j$. This ensures that \hat{p}_i^t values are nondecreasing. For each dose i , based on the samples q_i^t and \hat{p}_i^t , a corresponding utility score is $U^t(\hat{p}_i^t, q_i^t) = f_1(\hat{p}_i^t)f_2(q_i^t)$. Then, the estimated posterior expected utility is given by

$$\hat{E}[U(p_i, q_i) | D] = \frac{1}{T} \sum_{t=1}^T U^t(\hat{p}_i^t, q_i^t)$$

Finally, selected the optimal dose \hat{d} given by

$$\hat{d} = \operatorname{argmax}_i \hat{E}[U(p_i, q_i) | D]$$

Trial conduct

Trial conduct under TEPI is simple and transparent. During the trial, all dose-finding decisions follow Table 2 and the safety and futility rules. The steps of implementing TEPI design are as follows:

- Clinicians choose a starting dose, the maximum tolerable toxicity rate (p_T), and the minimum acceptable efficacy rate (q_E).
- Elicit the preset decision table (e.g., Table 1) and derive the dose-finding table (e.g., Table 2) to ensure that they reasonably reflect clinical practice during the trial (calibrate the intervals in the decision table as needed based on computer simulations).
- A dose is "available" if $Pr(p_i > p_T | D) < \eta$ and $Pr(q_i > q_E | D) > \xi$. If no dose is available, terminate the trial.
- The "current" dose is the dose used to treat the current cohort of patients.
- If the current dose invokes the safety rule, de-escalate to the closest available dose below the current dose.
- If the current dose invokes the futility rule,
 - If the decision is "E," escalate to the closest available dose above the current dose. If no doses above the current dose are available, de-escalate to the closest available dose below the current dose. (This rule is justified because we do not assume a monotonic dose–efficacy relationship. That is, although the current dose is not effective, an effective dose could be either a higher dose or a lower dose).
 - If the decision is "D," de-escalate to the closest available dose below the current dose. If no doses below the current dose are available, terminate the trial.
 - If the decision is "DU_T," mark the current dose and all higher doses as unavailable. De-escalate to the closest available dose below the current dose. If no doses are available, terminate the trial.
 - If the decision is "S," de-escalate to the closest available dose below the current dose. If no dose below the current dose is available, terminate the trial.
- If the current dose invokes both safety and futility rules, de-escalate to the closest available dose below the current dose. If no doses below the current dose are available, terminate the trial.
- Do not skip untried doses.
- Stop the trial at a prespecified sample size if it has not been terminated by then.
- Select the optimal dose \hat{d} in (2) as the final dose.

Operating Characteristics

Simulation setup

To characterize the TEPI design and compare the performance of TEPI with existing designs, we simulate clinical trials under the following six different scenarios. In most ACT dose-finding trials to date, four or fewer doses are investigated (12, 18–20). Therefore, we assume four doses in each scenario. For each scenario, we specify the true toxicity and efficacy probabilities for all four doses, and generate random binary toxicity and efficacy outcomes based on these probabilities. We compare with toxicity-based designs, such as the 3 + 3, mTPI, and CRM designs, and with toxicity–efficacy-based designs, such as the EffTox design (13). A full description of all the scenarios is provided in Supplementary Material C.

For TEPI, the simulation is based on the dose-finding decision described in Tables 1 and 2 using hyperparameters $\alpha_p = \beta_p = \alpha_q = \beta_q = 1$. For mTPI, the equivalence interval is set to [0.25, 0.35], and the dose-finding table is given in Table 3. We use 0.35 as the upper bound of the equivalence interval for mTPI instead of 0.4, because it is difficult to justify such a high-toxicity rate without considering efficacy data. NextGen-DF (4) is used to implement the standard 3 + 3 design and the CRM design (21). For the EffTox design, we use the EffTox software downloaded from https://biostatistics.mdanderson.org/software/download/SingleSoftware.aspx?Software_Id=2 and set $p_T = 0.4$, $q_E = 0.4$, $\pi_1 = (0.2, 0)$, $\pi_2 = (1, 0.6)$, and $\pi_3 = (0.5, 0.5)$, which is compatible with the proposed utility function. Under each simulation scenario, we ran 1,000 simulated trials with a maximum sample size of 27 patients and cohort size of 3.

Simulation results

Table 4 summarizes the simulation results. In scenario 1, all four doses are tolerable but unacceptable due to low efficacy rates. Under the TEPI design, 35% of the trials terminate early with an average sample size of 21. In contrast, the mTPI and CRM designs rarely stop early and exhaust the sample size, whereas the 3 + 3 design stops early 22% of the time and EffTox stops early 92% of the time, which means that EffTox is most efficient at minimizing the number of patients treated on a drug (or the selected dose ranges of a drug) having minimal efficacy. Except for EffTox, TEPI is better than other designs in this scenario.

Scenario 2 is an extreme case where all doses are safe but have the same high efficacy rate, so the starting dose has the highest utility. TEPI selects this dose level 84% of the time compared with 6% to 66% for the other designs. Besides, TEPI puts more patients

Table 3. Dose-finding decision table of the mTPI design

Number of DLTs	Number of patients at current dose									
	3	6	9	12	15	18	21	24	27	
0	E	E	E	E	E	E	E	E	E	
1	S	E	E	E	E	E	E	E	E	
2	D	S	S	E	E	E	E	E	E	
3	DU _T	S	S	S	S	E	E	E	E	
4		DU _T	S	S	S	S	E	E	E	
5		DU _T	DU _T	S	S	S	S	S	E	
6		DU _T	DU _T	D	S	S	S	S	S	
7			DU _T	DU _T	S	S	S	S	S	
8			DU _T	DU _T	DU _T	S	S	S	S	
9			DU _T	DU _T	DU _T	DU _T	S	S	S	
10				DU _T	DU _T	DU _T	DU _T	S	S	
11				DU _T	DU _T	DU _T	DU _T	DU _T	S	
12				DU _T	DU _T	DU _T	DU _T	DU _T	DU _T	

NOTE: Target toxicity rate = 30%; equivalence interval: [0.25, 0.35].

Li et al.

Table 4. Simulation results comparing the proposed TEPI, mTPI, 3 + 3, CRM, and EffTox designs

Scenario	Dose level	True probability		Selection probability (%)					Number of subjects treated				
		Tox	Eff	TEPI	mTPI	3 + 3	CRM	EffTox	TEPI	mTPI	3 + 3	CRM	EffTox
1	1	0.16	0.05	22.1	12.1	23.8	7.9	0	6.02	7.7	4.6	7.9	3.1
	2	0.2	0.1	17.9	24.1	22.0	23.6	0	5.7	8.1	3.8	7.5	3.1
	3	0.25	0.15	17.2	30.6	16.0	31.9	0	5.1	6.1	2.7	6.3	3.8
	4	0.3	0.18	7.5	32.1	15.8	36.6	8	4.3	4.9	1.4	5.2	5.3
					TEPI		mTPI		3 + 3		CRM		EffTox
		Probability of early termination			35.3		1.1		22.4		0.0		92
		Average number of subjects treated			21.2		26.8		12.6		27.0		15.3
2	1	0.15	0.8	83.9	10.7	23.4	6.3	66	9.1	7.4	4.5	7.5	17.8
	2	0.2	0.8	13.6	25.8	22.6	23.6	30	8.5	8.3	3.9	7.6	8.4
	3	0.25	0.8	2.1	28.2	17.1	32.7	3	5.6	6.0	2.9	6.5	0.7
	4	0.3	0.8	0.3	34.6	16.1	37.4	1	3.8	5.1	1.5	5.3	0.1
					TEPI		mTPI		3 + 3		CRM		EffTox
		Probability of early termination			0.1		0.0		20.8		0.0		0
		Average number of subjects treated			27.0		27.0		12.8		27.0		27.0
3	1	0.1	0.1	7.2	4.0	25.8	2.1	3	4.4	5.4	4.4	5.5	3.7
	2	0.2	0.7	88.0	32.5	36.4	32.1	42	12.3	9.8	4.6	8.9	11.8
	3	0.3	0.2	0.3	60.7	25.9	60.6	3	7.1	9.8	3.5	10.0	4.3
	4	0.7	0.1	0.1	2.5	1.4	5.2	2	2.3	1.9	1.1	2.6	2.0
					TEPI		mTPI		3 + 3		CRM		EffTox
		Probability of early termination			4.4		0.3		10.5		0.0		50
		Average number of subjects treated			26.1		26.9		13.6		27.0		21.8
4	1	0.15	0.43	53.9	11.3	23.8	7.7	19	6.1	7.3	4.6	7.7	7.3
	2	0.2	0.52	41.3	41.6	41.0	43.8	49	9.6	9.6	4.3	9.7	12.4
	3	0.4	0.5	3.6	38.8	11.7	41.7	22	9.0	7.9	2.8	7.6	5.3
	4	0.5	0.6	1.2	7.1	2.5	6.8	5	2.1	1.9	0.7	2.0	1.1
					TEPI		mTPI		3 + 3		CRM		EffTox
		Probability of early termination			1.2		1.2		21.0		0.0		6
		Average number of subjects treated			27.0		27.0		12.4		27.0		26.1
5	1	0.1	0.2	16.4	5.1	27.0	3.1	1	4.7	5.6	4.4	5.4	3.6
	2	0.2	0.6	65.4	31.5	34.8	25.0	48	8.6	9.6	4.6	8.2	12.4
	3	0.3	0.6	13.8	44.4	19.5	45.4	38	7.2	8.1	3.4	8.6	8.5
	4	0.4	0.6	1.0	19.0	9.1	26.5	10	4.9	3.7	1.4	4.8	2.0
					TEPI		mTPI		3 + 3		CRM		EffTox
		Probability of early termination			3.4		0.0		9.6		0.0		3
		Average number of subjects treated			26.3		27.0		13.8		27.0		26.5
6	1	0.5	0.4	33.9	23.0	10.7	99.3	16	14.9	14.0	4.5	25.8	7.9
	2	0.6	0.5	0.3	0.9	0.3	0.7	13	1.8	1.2	0.6	1.1	6.6
	3	0.7	0.6	0.0	0.1	0.2	0.0	2	0.1	0.1	0.0	0.1	1.2
	4	0.8	0.8	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.2
					TEPI		mTPI		3 + 3		CRM		EffTox
		Probability of early termination			65.8		76.0		88.8		0.0		69
		Average number of subjects treated			16.8		15.3		5.2		27.0		15.9

NOTE: If one exists, the dose level with the best utility is in boldface.

Abbreviations: Eff, efficacy; Tox, toxicity.

on average at dose levels 1 and 2 and fewer patients at dose levels 3 and 4 than the mTPI and CRM designs.

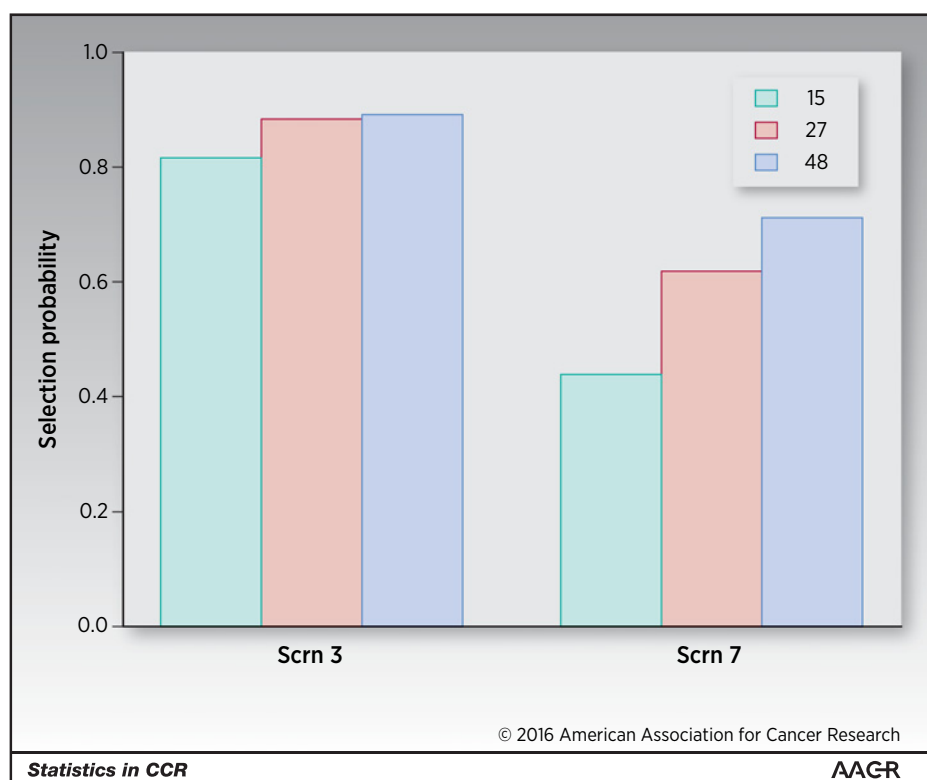
Under scenario 3, dose levels 1 to 3 are safe, dose level 4 is unsafe, and dose level 2 has the highest efficacy. TEPI selects dose level 2 88% of the time and allocates an average of 12 patients to it. In contrast, mTPI and CRM select this dose level with much lower frequencies and allocate fewer patients. Moreover, mTPI and CRM select dose level 3 (the MTD) 60% of the time, whereas TEPI selects

this level only 0.3% of the time. A similar trend is observed for dose level 4. While EffTox selects dose level 2 mostly in trials that do not stop early, it stops early with a rate as high as 50%, which is not necessary because there is a desirable dose in this scenario.

In scenario 4, the two lower doses are safe and effective, while the two higher doses are efficacious but unsafe. There is a monotonically increasing relationship between efficacy and dose, which is the underlying assumption for the mTPI, 3 + 3, and CRM

Figure 2.

Relationship between sample size and the frequency of selecting the optimal dose in two arbitrarily selected scenarios (Scrn 3 and 7). Results for other scenarios are similar, but not shown.



designs. In this scenario, dose level 2 has the highest utility. The four designs select this dose with similar frequency (~40%), which demonstrates that TEPI has good performance characteristics even under the conventional assumptions. However, it is interesting to note that 54% of the time, TEPI selects dose level 1, which has slightly lower utility than dose level 2.

On the other hand, the mTPI, CRM, and EffTox designs are aggressive, selecting an unsafe dose (level 3 or 4) 46%, 49%, and 27% of the time, respectively; in comparison, the 3 + 3 design selects an unsafe dose 14% of the time, and TEPI only 5% of the time.

In scenario 5, all doses are tolerable, and efficacy increases from dose level 1 but plateaus at dose level 2. Dose level 2 is optimal. TEPI selects this dose level 65% of the time, allocating an average of 8.6 patients. In contrast, mTPI, CRM, and EffTox select dose level 2 32%, 25%, and 48% of the time, respectively, while selecting the suboptimal dose level 3 44%, 45%, and 38% of the time.

In scenario 6, all doses are too toxic despite having acceptable efficacy. In this case, TEPI terminates early 66% of the time, with an average sample size of 17, and mTPI, 3 + 3, and EffTox terminate early 76%, 89%, and 69% of the time, respectively, with average sample sizes of 15, 5, and 16. CRM does not stop early, with an average of 26 patients treated at dose levels 1 and 2. Despite its ability to stop early with a slightly higher rate than TEPI, EffTox aggressively selects dose level 2 12% of the time. Here, TEPI seems more aggressive than mTPI, which is expected because dose level 1 has acceptable efficacy.

In all scenarios where an acceptable dose exists (scenarios 2, 3, 4, 5), the 3 + 3 design is less likely to select the desirable dose compared with the TEPI. It appears that the 3 + 3 design is too conservative in that it is unable to escalate quickly, even when the doses are safe, consistent with previous finding (3). As CRM and mTPI do not incorporate efficacy data in the trial conduct, TEPI is

superior in scenarios 1 to 3 and 5 and performs well in scenario 4. EffTox performs better than TEPI only in scenario 1, and in scenarios 2 to 6, TEPI is more desirable than EffTox.

Sensitivity to sample size

We performed a sensitivity study to evaluate the impact of varying sample sizes on the ability of the TEPI design to identify the optimal dose. We arbitrarily selected scenario 3 and created a new scenario (scenario 7) with $p = (0.1, 0.2, 0.3, 0.7)$ and $q = (0.05, 0.2, 0.5, 0.6)$. Figure 2 plots the frequency of selecting the true optimal dose level against sample sizes of 15, 27, and 48. For each scenario and sample-size combination, we simulated 1,000 trials. In both scenarios, the selection probability increases with sample size. This observation is not surprising and supports the importance of considering the tradeoff between sample size and the precision of dose selection. However, it is worth noting that even for the small sample size of 15 or 27, the TEPI design performs reasonably well.

Discussion

Traditional phase I designs use only toxicity data and may not be appropriate for ACT dose-finding trials, because efficacy may not increase with increasing dose. The proposed TEPI design attempts to address this problem by accounting for both efficacy and toxicity simultaneously. Typically in ACT trials, efficacy or activity biomarkers can be observed quickly, oftentimes as fast as the toxicity outcomes. This makes TEPI feasible. However, when efficacy outcome cannot be quickly observed, either delayed enrollment is required to use TEPI or statistical models must be modified to account for delayed efficacy outcomes.

The TEPI design is simple, transparent, and appealing because all dose-escalation decisions can be prespecified prior to the trial start. During the trial conduct, no design modification or "black-box" statistical calculation is required. Through simulation studies, we have demonstrated the superiority of the TEPI design in selecting and allocating more patients to the true optimal dose over the 3 + 3, mTPI, and CRM designs, especially when the assumption of monotone increasing efficacy is violated. Even if the monotonic relationship is true, the proposed design is still superior to the 3 + 3 design. More importantly, TEPI terminates the trial earlier when all tested doses are safe but no dose is likely to be efficacious. This is certainly more ethical than to continue to expose patients with cancer to a safe but ineffective drug.

The proposed TEPI design is a natural extension of mTPI by adding the efficacy interval into the dose-finding model. To use this design properly, we recommend a close collaboration between clinicians and statisticians to determine the initial design parameters, such as the interval combinations and corresponding dose-finding decisions in Table 1, the utility function, and the safety and futility stopping rules.

The TEPI design uses the utility function of safety and efficacy to choose the optimal dose. Depending on the clinical situation, one could use a different metric to make the final dose selection. For example, the dose with the highest probability $Pr(p_i < p_T, q_i > q_E + \delta|D)$ may be selected as the best dose, where δ is the expected increment over the minimum efficacy rate. A

limitation of TEPI may be the assumption of independence of the safety and efficacy data. In general, the true relationship between safety and efficacy is complex, and it has been demonstrated that this independence assumption has negligible effects (22, 23). Finally, we reemphasize the importance of simplicity, flexibility, and transparency of using TEPI in designing a dose-finding trial.

Disclosure of Potential Conflicts of Interest

D.H. Li holds equity interest (including patents) in Juno Therapeutics, Inc. J.B. Whitmore holds equity interest in Juno Therapeutics, Inc. Y. Ji is co-founder of and holds ownership interest (including patents) in Laiyi Consulting, Inc. and is a consultant for Takeda Pharmaceuticals USA, Inc. No potential conflicts of interest were disclosed by the other author.

Authors' Contributions

Conception and design: D.H. Li, J.B. Whitmore, W. Guo, Y. Ji
Development of methodology: D.H. Li, J.B. Whitmore, W. Guo, Y. Ji
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): D.H. Li, J.B. Whitmore, W. Guo, Y. Ji
Writing, review, and/or revision of the manuscript: D.H. Li, J.B. Whitmore, W. Guo, Y. Ji
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D.H. Li

Received May 4, 2016; revised July 15, 2016; accepted August 12, 2016; published OnlineFirst October 14, 2016.

References

1. Storer BE. Design and analysis of phase I clinical trials. *Biometrics* 1989;45:925-37.
2. Ji Y, Liu P, Li Y, Bekele BN. A modified toxicity probability interval method for dose-finding trials. *Clin Trials* 2010;7:653-63.
3. 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. *J Clin Oncol* 2013;31:1785-91.
4. Yang S, Wang SJ, Ji Y. An integrated dose-finding tool for phase I trials in oncology. *Contemp Clin Trials* 2015;45:426-34.
5. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990;46:33-48.
6. Davila ML, Brentjens R, Wang X, Riviere I, Sadelain M. How do CARs work? Early insights from recent clinical studies targeting CD19. *Oncoimmunology* 2012;1:1577-83.
7. Kochenderfer JN, Yu Z, Frasher D, Restifo NP, Rosenberg SA. Adoptive transfer of syngeneic T cells transduced with a chimeric antigen receptor that recognizes murine CD19 can eradicate lymphoma and normal B cells. *Blood* 2010;116:3875-86.
8. Porter DL, Levine BL, Kalos M, Bagg A, June CH. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N Engl J Med* 2011;365:725-33.
9. Johnson LA, Morgan RA, Dudley ME, Cassard L, Yang JC, Hughes MS, et al. Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen. *Blood* 2009;114:535-46.
10. Husain S, Han J, Au P, Shannon K, Puri R. Gene therapy for cancer: regulatory considerations for approval. *Cancer Gene Ther* 2015;22:554-63.
11. Park JH, Riviere I, Wang X, Bernal Y, Purdon T, Halton E, et al. Efficacy and safety of CD19-targeted 19-28z CAR modified T cells in adult patients with relapsed or refractory B-ALL. In: ASCO Annual Meeting Proceedings. 2015;33:7010.
12. Gardner RA, Park JR, Kelly-Spratt KS, Finney O, Smithers H, Hoglund V, et al. T Cell Products of Defined CD4:CD8 Composition and Prescribed Levels of CD19-CAR/EGFR Transgene Expression Mediate Regression of Acute Lymphoblastic Leukemia in the Setting of Post-Allo- HSCT Relapse; 2014. Presented at the 56th Annual Meeting of the American Society of Hematology, San Francisco, CA.
13. Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 2004;60:684-93.
14. Ji Y, Li Y, Bekele BN. Dose-finding in phase I clinical trials based on toxicity probability intervals. *Clin Trials* 2007;4:235-44.
15. Thall PF, Nguyen HQ. Adaptive randomization to improve utility-based dose-finding with bivariate ordinal outcomes. *J Biopharm Stat* 2012;22:785-801.
16. Lee J, Thall PF, Ji Y, Müller P. Bayesian dose-finding in two treatment cycles based on the joint utility of efficacy and toxicity. *J Am Stat Assoc* 2015;110:711-22.
17. Quintana M, Li DH, Albertson TM, Connor JT. A Bayesian adaptive phase 1 design to determine the optimal dose and schedule of an adoptive T-cell therapy in a mixed patient population. *Contemp Clin Trials* 2016;48:153-65.
18. Sauter CS, Riviere I, Bernal Y, Wang X, Purdon T, Yoo S, et al. Phase I trial of 19-28z chimeric antigen receptor modified T cells (19-28z CAR-T) post-high dose therapy and autologous stem cell transplant (HDT-ASCT) for relapsed and refractory (rel/ref) aggressive B-cell non-Hodgkin lymphoma (B-NHL). In: ASCO Annual Meeting Proceedings. 2015;33:8515.
19. Turtle CJ, Hanafi LA, Berger C, Gooley TA, Cherian S, Hudecek M, et al. CD19 CAR-T cells of defined CD4+: CD8+ composition in adult B cell ALL patients. *J Clin Invest* 2016;126:2123-38.
20. Fry TJ. CD22 CAR update and novel mechanisms of leukemic resistance; 2016. Presented at the 2016 AACR, New Orleans, LA.
21. Cheung Y. dcfcm: Dose-finding by the continual reassessment method. R package version. 2013; p. 02-2.
22. Cai C, Yuan Y, Ji Y. A Bayesian dose finding design for oncology clinical trials of combinational biological agents. *J R Stat Soc Ser C Appl Stat* 2014;63:159-73.
23. Guo W, Ni Y, Ji Y. TEAMS: Toxicity-and efficacy-based dose-insertion design with adaptive model selection for phase I/II dose-escalation trials in oncology. *Stat Biosci* 2015;7:432-59.

Clinical Cancer Research

Toxicity and Efficacy Probability Interval Design for Phase I Adoptive Cell Therapy Dose-Finding Clinical Trials

Daniel H. Li, James B. Whitmore, Wentian Guo, et al.

Clin Cancer Res 2017;23:13-20. Published OnlineFirst October 14, 2016.

Updated version Access the most recent version of this article at:
[doi:10.1158/1078-0432.CCR-16-1125](https://doi.org/10.1158/1078-0432.CCR-16-1125)

Supplementary Material Access the most recent supplemental material at:
<http://clincancerres.aacrjournals.org/content/suppl/2016/10/14/1078-0432.CCR-16-1125.DC1>

Cited articles This article cites 20 articles, 3 of which you can access for free at:
<http://clincancerres.aacrjournals.org/content/23/1/13.full#ref-list-1>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://clincancerres.aacrjournals.org/content/23/1/13>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.