Published in final edited form as:

Stat Biopharm Res. 2018 ; 10(3): 185–195. doi:10.1080/19466315.2018.1462727.

An adaptive dose-finding design based on both safety and immunologic responses in cancer clinical trials

Cody Chiuzan^a, Elizabeth Garrett-Mayer^b, and Michael Nishimura^c

^aDepartment of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, USA.

^bCenter for Research and Analytics (CENTRA), American Society of Clinical Oncology, VA, USA.

^cDepartment of Surgery, Cardinal Bernardin Cancer Center, Loyola University, Chicago, Maywood, IL, USA.

Abstract

Dose-finding in cancer clinical trials has been dominated by algorithmic designs on the principle that the highest tolerable dose is also the most effective dose. This assumption no longer applies to the biologic treatments that are characterized by different toxicity and/or efficacy profiles to the extent that the best therapeutic dose might be well below any dose that produces serious toxicity. As such, we propose a two-stage design with focus on immunotherapy trials, incorporating both safety and efficacy information. The 1st stage establishes the safety profile of each dose, with escalation decisions based on likelihood principles. Continuous immunologic outcomes are used to evaluate the relative efficacy of the doses. The 2nd stage employs an adaptive randomization to assign patients to doses showing higher efficacy. Safety is being continuously monitored throughout stage 2, where some doses may be ‘closed’ due to unacceptable toxicity. The proposed design is compared to the modified toxicity probability interval (mTPI) design using percent dose allocation and estimation of outcomes under different scenarios. We show that by using an efficacy-driven adaptive randomization with safety constraints, the allocation distribution is skewed towards more efficacious doses, and thus limit the number of patients exposed to toxic or non-therapeutic doses.

Keywords

Adaptive design; cancer immunotherapies; phase I trials

1 Introduction

1.1 Motivation

Clinical cancer research has made remarkable advances in methods of treatment over the past decade. Cytotoxic drugs continue to be the predominant part of most therapies, but innovative agents addressing molecular targets are being incorporated more and more

frequently into the standard of care. The main assumption driving the design of phase I cytotoxic trials has been that higher doses are associated with higher toxicity, but also higher efficacy, and so oncology drug development dose finding has sought the highest tolerable doses for anti-cancer treatment candidates. Specifically, the main goal of phase I trials for cytotoxic agents is to identify the maximum tolerated dose (MTD) and pharmacokinetics (PK) in order to determine dose-safety and dosing-schedule for subsequent phase II and III. However, the new trend of targeted and biologic agents, does not necessarily adhere to the principles of the classical phase I designs. In an immunotherapy setting, the MTD is not necessarily the most effective dose to induce anti-tumor activity and some treatment types do not provide any PK data [Herberman 1985].

Cancer immunotherapeutics, including therapeutic cancer vaccines, monoclonal antibodies, and therapeutic proteins have emerged as promising treatment modalities in part because of their minimal toxicities [Roberts et al. 2004, Horstmann et al. 2005] and early signs of benefit such as slowing tumor growth. The treatment methods are designed specifically to either 1) stimulate the body's immune system to recognize and attack cancer cells, or 2) provide immune system components, such as proteins (antibodies), meant to kill tumor cells either by direct action, induction of apoptosis, or delivery of cytotoxic drug/agent. On these premises, it is assumed that the dose-toxicity curve is shallower and that efficacy does not necessarily increase monotonically with higher doses. The immune response (an indicator of efficacy) might very well plateau after reaching a maximum, and therefore any dose past that point would have no additional impact on efficacy. In addition, more recent early trials incorporate a proof-of-principle of biologic effect, i.e., evidence of antitumor activity or other immunologic parameters, and target the optimal biological dose rather than the MTD. Given the relatively acceptable safety profile of immunotherapies, toxicity and efficacy should be explored together to maximize the number of patients allocated to more therapeutic doses and perhaps accelerate the dose selection process.

1.2 Early Phase Designs in Cancer Trials

The majority of the dose-finding designs implemented in oncology trials are focused on dose-safety alone based on binary or ordinal toxicity outcomes [Van Meter, Garrett-Mayer and Bandyopadhyay 2012, Yuan, Chappell and Bailey 2007]. Statistical approaches vary from the standard algorithmic '3+3' [Storer 1989] to more sophisticated model-based designs such as the continual reassessment method (CRM) [O'Quigley, Pepe and Fisher 1990] and its variations [Piantadosi, Fisher and Grossman 1998, Moller 1995, Cheung and Chappell 2000]. In the last decades, the accelerating therapeutic development of non-cytotoxic agents motivated the inclusion of additional endpoints such as efficacy, pharmacokinetics (PK), or pharmacodynamics (PD). Many studies have used a '3+3' with an expansion cohort (EC) to increase the patient allocation at the MTD in order to gather additional information concerning toxicity or include efficacy data. While there is no design specific to EC, the method continues to be used with no statistical justification of the sample size that dramatically changed in size within the last decade [Dahlberg et al. 2014]. When toxicity and efficacy are correlated, seamless phase I/II designs offer multiple possibilities for trial implementation. Their goal is to select a dose with high efficacy among dose levels with acceptable toxicity, or the dose with the optimal efficacy-toxicity profile. Most of these

designs are Bayesian methods and take one of two approaches: 1) a decision criterion is based on trade-offs between probabilities of treatment efficacy and toxicity [Thall and Cook 2004, Thall, Cook and Estey 2006, Thall and Russell 1998], and 2) the use of elicited joint utilities of ordinal (efficacy, toxicity) outcomes [Yuan, Chappell and Bailey 2007, Braun 2002, Zhang, Sargent and Mandrekar 2006].

For immunotherapy trials, Hoering et al. 2011 and Messer et al. 2010 proposed two phase I/II designs for targeted agents and cancer vaccines, respectively. Both approaches implement a conventional standard design (e.g., '3+3') for testing safety and finding the MTD, followed by an efficacy stage. Hoering et al. employ a modified phase II selection for two or three doses at and below the MTD through a correlated bivariate logistic regression model. Messer et al. proceed with a primary phase II addressing an efficacy point (i.e., progression free survival at one year).

Compared to the traditional methods, these novel designs show attractive statistical properties in terms of sample size reduction, improved precision of estimates, and skewed allocation distribution towards more promising/effective doses. Nevertheless, clinical researchers continue to be reluctant to implement new methods, often because of their technical and computational aspects. Additionally, cancer immunotherapy trials face challenges of choosing and incorporating meaningful immunological outcomes into the dose selection process and defining optimization criteria for the case where no DLT (dose limiting toxicity) is expected or achieved.

We propose an adaptive dose-finding design for assessing toxicity and continuous efficacy outcomes in cancer immunotherapy trials. The goal is to select the dose that optimizes some efficacy criterion while maintaining safety. Although it is often assumed that cancer immunotherapies are in general safe or have low toxicity, in the initial stage we establish the safety of each dose. Fixed cohorts of patients are assigned sequentially to prespecified doses, with dose escalation decisions based on likelihood inferences. If two or more doses have been considered acceptable, in the second stage we use an efficacy-driven adaptive randomization to allocate the remaining patients. Safety data is collected throughout stage 2. When a DLT is observed, we employ the likelihood method (defined in stage 1) to decide if a dose continues to demonstrate adequate safety and if it should still be considered for allocation. Comparisons with the modified toxicity probability interval (mTPI) design [Ji and Wang 2013] are made in terms of number of patients treated per dose and average sample size.

The design imposes monotonicity on the dose-toxicity relationship, but it is flexible in allowing a non-monotone dose-efficacy curve. Fit in frequentist framework, the model is easy to implement and it can be tailored to different toxicity and activity profiles, especially in testing immune-targeted agents for which identifying the MTD might be less relevant in selecting the recommended phase 2 dose (RP2D). Most importantly, it offers a probabilistic alternative to any of the rule-based algorithms. R functions with documentation for implementing the design and performing simulations in the planning stage of a trial are available upon request.

2 Methods

2.1 Evidential Paradigm

The first stage of our proposed model falls within the evidential paradigm framework. The concept was first introduced by Hacking in 1965 by stating the formal expression of the Law of Likelihood [Hacking 1965]. The evidential paradigm requires the specification of two simple hypotheses (treated symmetrically) and uses the likelihood-ratio (LR) as an objective measure of the strength of evidence decoupled from the uncertainty (probability of observing weak or misleading evidence). Given two hypotheses $H_1: \theta = \theta_1$ and $H_2: \theta = \theta_2$, strong evidence supporting θ_2 over θ_1 exists if $LR = L(\theta_2; x)/L(\theta_1; x) \geq k$. Weak evidence exists if $1/k < LR < k$, and misleading evidence exists when $LR \leq k$ under H_1 true.

Compared to the Neyman-Pearson approach that requires the specification of type I and II errors, the evidential paradigm involves only criterion k for specifying the probability of strong evidence. Royall and Blume established a correspondence between the values of k and type I and II errors [Royall 1997, Blume 2002]. They proposed benchmarks of 8 and 32, representing ‘weak’ ($1 < LR < 8$), ‘fairly strong’ ($LR \geq 8$), and ‘strong’ ($LR \geq 32$) levels of evidence. In early phase trials, control of the error rates is usually less stringent. However, an extremely useful feature for sequential trials is that the probability of misleading evidence under the true H_1 hypothesis (analogous to the classical type I error) is bounded, i.e., cannot exceed $1/k$ regardless of the number of repeated looks [Royall 1997, Royall 2000].

2.2 Two-Stage Adaptive Model

Stage 1: Establish safety and collect information on immunological outcomes.

Cohorts of m_j patients are assigned sequentially to increasing doses, with escalation not occurring until acceptability has been determined in the previous cohort. Even though dose escalation proceeds in an algorithmic fashion, the decision-making process has a likelihood framework.

Let $d_j, j = 1, 2, \dots, D$, be the set of ordered dose levels. For each subject assigned to a dose d_j , a single binary variable T indicates whether the patient has experienced a DLT: $T = 1$ with probability p_j (yes DLT), and $T = 0$ with probability $1 - p_j$ (no DLT), where p_j is the true unknown DLT rate at each dose. Let p_1 and p_2 be the unacceptably high and acceptable toxicity rates, respectively. For any p_1 and p_2 established *a priori*, the following hypotheses are to be tested for each dose level:

$$H_1: p_j = p_1 \text{ vs } H_2: p_j = p_2.$$

Suppose that among the m_j subjects equally allocated to each dose, t_j have experienced DLTs. The likelihood-ratio (LR_j) for each dose given the observed data is:

$$LR_j = \left(\frac{p_2}{p_1} \right)^{t_j} \left(\frac{(1-p_2)}{(1-p_1)} \right)^{m_j - t_j}, j = 1, 2, \dots, D.$$

Based on LR_j and a certain benchmark k , the strength of evidence is classified in one of the following three categories:

1. Weak evidence (supporting neither hypothesis), if $1/k < LR_j < k$
2. Evidence supporting H_2 (acceptable dose), if $LR_j \geq k$
3. Evidence supporting H_1 (unsafe dose), if $LR_j \leq 1/k$.

An important aspect of the study is the choice of k representing the level of evidence. Under an assumed DLT rate, the following operating characteristics can be computed for a range of k values: 1) probability of weak evidence, 2) probability of correctly declaring a dose acceptable, and 3) probability of correctly declaring a dose unsafe. The first category is ambiguous and can be subject to interpretability. In our model, we regard it as not having enough information to conclude toxicity and combine it with evidence supporting H_2 . That is, a dose is considered safe until there is sufficient evidence accumulated to conclude otherwise. For any k there is a trade-off between the probabilities of correctly identifying an acceptable and an unsafe dose. For example, in the case of cancer immunologic agents that have tolerable safety profiles, we would aim for a higher probability of correctly declaring a dose acceptable at the expense of a lower probability of identifying an unsafe dose. Thus, selecting k as a threshold for the likelihood-ratio is not a one-size-fits-all system, but rather a context-specific problem.

Once a threshold has been chosen, safety is determined using the following rule: if $LR_j > 1/k$, $j = 1, 2, \dots, D$, dose j is regarded as acceptably safe and escalation to the next higher dose is permitted; if $LR_j \leq 1/k$, dose j is thought to be unacceptably toxic, the first stage is stopped, and all acceptable doses are carried forward to stage 2. It is expected that due to the limited sample size at this stage, the likelihood-ratios would have values relatively close to 1, and therefore a cutoff of $1/32$ would rarely be crossed. Under these circumstances, we recommend $k = 8$ as the maximum value for classifying evidence. Actually, it has been proposed that $k = 2$ is a reasonable threshold for phase I studies, which is consistent with the level of evidence historically accepted in the '3+3' design [Chiuzan, Garrett-Mayer, and Yeatts 2015]. The general setting of this likelihood method allows for different combinations of testing hypotheses, cohort sizes per dose, k values, and true toxicity rates. In section 3 we illustrate its implementation for cohorts of 3 patients per dose and compare it to the modified probability interval (mTPI) design, a mixture of rule-based and model-based methods with superior properties as compared to those of the '3+3' algorithm [Ji and Wang 2013].

Stage 2: Employ an efficacy-driven adaptive randomization based on continuous efficacy endpoints.

The continuous nature of the outcomes draws flexibility of incorporating endpoints from various research areas. Examples of continuous efficacy outcomes that have been used in phase I studies include: percent persistence of T-cells in immunology, fold-change for genetic markers, or absolute count in pharmacokinetics.

Once stage 1 has been completed, all acceptable doses are taken to stage 2, where the remaining patients are adaptively randomized to these doses based on efficacy outcomes. Let

N be the maximum sample size, and d_j be the set of acceptable doses after stage 1, with $1 < j \leq D$. If $j = 1$, that is, only one dose is considered acceptable after stage 1, then there is no need for randomization.

For each of the acceptable doses, we define Y_{ij} as the efficacy outcome collected from patient i assigned to dose d_j . A linear model is fit to the efficacy data collected in stage 1 in order to obtain the efficacy estimates associated with each dose level:

$$Y_{ij} | d_j = \beta_0 + \sum_{j=1}^{D-1} \beta_j d_j + e_{ij}; i \in \{1, 2, \dots, N\}, 1 < j \leq D.$$

For situations where the normality assumption is problematic, different transformations of the efficacy outcome can be used as a remedy, e.g., natural logarithm, square-root, etc. The model is only linear in parameters β , and thus it can accommodate different forms of the dose variables. For example, dose assignment can be regarded as a categorical predictor and modeled using indicator (dummy) variables, or it can be considered of continuous nature and estimated with linear and/or quadratic terms. After estimating the efficacy outcomes, randomization probabilities are calculated for all acceptable dose levels:

$$\pi_j = \hat{Y}_j \left/ \sum_{r=1}^D \hat{Y}_r \right., 1 < j \leq D.$$

The next patient will be randomized to dose d_j with probability π_j , where doses with higher efficacy are more likely to be assigned. If a DLT is observed, the dose safety is re-evaluated using the likelihood approach from stage 1. As data becomes available, the model is refitted and new randomization probabilities are being generated for accruing patients. The process is repeated until the maximum sample size N is achieved or some other predefined stopping rule is met.

Although the second stage is efficacy driven, safety is being continuously monitored. For every DLT occurrence we employ the likelihood method described in stage 1 using all available toxicity data at that dose. Due to the monotonicity assumption of the dose-toxicity association, an unacceptably toxic dose and all higher doses will be considered ‘closed’ and excluded permanently from randomization. For situations when only one dose remains safe, we continue allocation until a predefined maximum number of patients (usually between 9 and 20) have been treated at that level. This expansion allows us to obtain a more precise estimate of toxicity and efficacy for the most promising doses before initiating a phase II trial with clinical outcomes. The trial will stop early if no dose is considered safe.

3 Simulation Studies

3.1 Assumptions and Simulation Models

We consider a binary toxicity measure and a continuous efficacy outcome. The two endpoints are modeled independently, based on the assumption that toxicity does not affect the possibility of observing efficacy. The only monotonicity constraint is imposed on the

dose-toxicity relation, while the dose-efficacy curve is allowed to be non-monotonic. The efficacy outcome chosen for demonstration is inspired by an ongoing adoptive T-cell therapy (ACT) trial in metastatic melanoma patients (NUT P01 CA154778-01). It represents percent persistence of transduced (genetically modified to display an antigen specific receptor) T-cells at 30 days relative to baseline. In our simulations, both toxicity and efficacy outcomes are generated at the individual level, with persistence (%) data assumed to follow a beta-binomial distribution, given dose:

$$Y_{ij} \sim \text{Bin}(100, \mu_j), i \in \{1, 2, \dots, N\}, 1 < j \leq D,$$

$$\mu_j \sim \text{Beta}(\alpha, \beta).$$

The within-patient variability is controlled by the binomial parameters, whereas between-patient variability at each dose is controlled by the beta distribution, and it can vary or be assumed constant across doses. In this paper we apply a natural log transformation for the efficacy outcomes and use indicator variables for modeling dose levels, which allows maximum flexibility for the dose-efficacy relationship.

The example chosen for illustration uses 5 dose levels and two total sample sizes: $N = 25$ and $N = 50$, with cohorts of $m_j = 3$ patients assigned in stage 1 to each of the predefined doses $d_j, j = 1, 2, 3, 4, 5$. In simulations, we assumed that persistence variability is constant across dose levels, but tested two different values: 0.2% and 1%. In terms of persistence outcomes, using a variance of 0.2% would ensure that 95% of the values vary within $\pm 9\%$ from the mean, compared to $\pm 20\%$ when applying a variance of 1%. Illustrations of these two variability scenarios are shown in Figure 1 for a total sample size of 25 patients and a plateau trend for T-cell persistence as function of dose. We chose to report only the results associated with the largest patient heterogeneity. The operating characteristics of the likelihood method implemented in stage 1 have already been studied for several sets of hypotheses (p_1, p_2) and k thresholds [Chiuzan, Garrett-Mayer and Yeatts 2015]. For illustration purposes, in this paper, the model behavior is presented for one pair of hypotheses: $p_1 = 0.40$ (unacceptably high toxicity), $p_2 = 0.15$ (acceptable toxicity) and threshold $k = 2$. Extensions and implementations to other hypotheses follow similar steps and can be tailored based on the characteristics of the study agent.

Four toxicity and four efficacy scenarios are evaluated with 5,000 simulated trials per each combination. Figure 2A depicts the efficacy trends as functions of dose levels. Eff1 scenario assumes a linear increase in T-cell persistence with increasing dose level. Eff2 scenario considers a non-linear (umbrella) trend where persistence reaches the maximum at dose level 3 and then decreases for doses 4 and 5. In Eff3 scenario we assume an increase in T-cell persistence up to dose 3 after which it plateaus. Eff4 scenario describes a T-cell percent persistence among all dose levels. Figure 2B displays the toxicity probabilities as functions of dose levels. Tox1 and Tox4 assume constant toxicities across dose levels: high and moderate DLT rates, respectively. Scenarios Tox2 and Tox3 show a monotonically increasing toxicity with dose, where Tox2 is steeper than Tox3.

In each scenario, the optimal dose is the dose producing the highest percent persistence of transduced T-cells at follow-up among all acceptable doses. The design is evaluated using the following operating characteristics: 1) the number of patients allocated to each dose, overall (Table 1) and separately by stage (Table S1 in Supplementary Materials), and 2) estimated persistence per dose level at the end of stage 2 (Table 2).

3.2 Model Comparison (mTPI)

We further compare our results with the modified toxicity probability interval (mTPI) design. Implemented in a Bayesian framework, the mTPI design uses posterior toxicity probabilities to determine the dose-escalation rules (intervals): ‘underdosing – escalate to the next dose level, ‘proper dosing’ – stay at the current dose, and ‘overdosing’ – de-escalate the dose. Compared to other rule-based methods (e.g., ‘3+3’), the mTPI design has demonstrated superior properties such as fewer patients treated above the MTD and higher probabilities of correctly selecting the true MTD. Additionally, Chiuzan et al (2015) have already illustrated the operating characteristics of the likelihood approach from stage 1 to the mTPI design, thus a comparison of the two models seems appropriate and relevant. The mTPI simulations use a target toxicity probability of 0.275 (the midpoint of the two hypotheses 0.40 vs 0.15), cohorts of three patients per dose with allocation starting at dose level 1. The two designs are compared in terms of percent allocation per dose for each toxicity/efficacy combination. In addition, for the mTPI design we report the average total sample sizes and probabilities of selecting the true MTD in Table S2, Supplementary Materials.

4 Results

Table 1 shows the allocation percentages per each dose level for all toxicity/efficacy scenarios. Results are summarized as median percentage and interquartile range (IQR) and presented from the highest (Tox1) to the lowest toxicity scenario (Tox4), for all efficacy trends (Eff1-Eff4). The allocation percentages per dose are calculated only out of the total of allocated patients. Specifically, for high toxicity scenarios (Tox1) the trial will terminate early with a maximum 50% allocation of the proposed sample size. Contrarily, for low toxicities (Tox4) the goal is to use the entire sample size and skew the allocation distribution towards acceptable doses with higher persistence.

In stage 1, we chose $p_1 = 0.40$ as the unacceptably high toxicity hypothesis, and $p_2 = 0.15$ as the acceptable toxicity value. The midpoint between these values 0.275 is used as the target toxicity probability for the mTPI comparison. Thus, we can reasonably use 25% as a cutoff for assessing which dose should be declared acceptable. For example, under Tox1 (0.35, 0.35, 0.35, 0.35, 0.35) there should be no acceptable doses, whereas under Tox4 (0.02, 0.02, 0.02, 0.02, 0.02) all doses should be considered tolerable, and the adaptive randomization skews the allocation towards levels with higher persistence. Specifically, for Tox3 scenario, doses 1, 2, and 3 are considered acceptably safe. Amongst these, dose level 3 has the highest persistence and should have the maximum percent allocation. Surely, as seen in Table 1, the design correctly assigns most of the patients to dose 3 for all combinations: Tox3/Eff1, Tox3/Eff2, and Tox3/Eff3. For a total sample size of 25 patients, the median allocation

ranges from 24% to 29% depending on the toxicity/efficacy scenario, and up to approximately 10% higher when the sample size doubles.

The most representative behavior in terms of allocating patients to more promising doses is under Tox4 that displays a constant toxicity probability of 2% across doses. Almost 30% of the patients are assigned to the highest persistence dose (dose 5 for Eff1, and dose 3 for Eff2), or have an equal allocation of 24% (IQR: 20-28%) at doses 3, 4, and 5 for the efficacy plateau trend Eff3. When there is no dose-efficacy relationship (flat trend Eff4), the design resembles the balanced model by allocating a median of 20% (IQR: 16-24%) of patients to each of the 5 doses.

As expected, if there are no acceptable doses (Tox1), the two-stage adaptive design stops allocating patients beyond dose 3, with the majority of treated patients randomized to one of the two lowest doses. In fact, in 75% of the trials up to 88% ($N=25$) and 94% ($N=50$) of the patients are not randomized to any doses considered unsafe.

Table S1 displays the number of patients included at each stage. For highly toxic scenarios (Tox1), the median number of patients at stage 1 does not exceed 9, with a constant median of 6 patients randomized at stage 2 (similar results for $N=25$ and $N=50$). In scenarios with 2 or more acceptable doses, the median allocation at stage 1 is of 15 patients, with the remaining sample size (10 or 35) being adaptively allocated in stage 2.

Table 2 shows the estimated efficacy per dose for all scenarios. The results are summarized as median of estimated percent persistence along with the corresponding IQR. For each scenario we also report the true mean efficacy per dose used for generating data. Results are compared in terms of precision and variability of estimated persistence. There is very little difference (approximately 0.5%) between the median estimates of persistence using 25 versus 50 patients. This could very well be an effect of using the median with its known robustness to outliers. In general, when the percent allocation per dose is 15% or less, the estimated median persistence is between 2-3% lower than the true mean persistence. When 20% or more patients are assigned to a certain dose, the estimated median persistence either equals the true value or lies within 0.5% difference. A similar pattern can be seen for differences in IQR. When a certain dose has roughly 15% of patients allocated, the IQRs generated by 25 and 50 patients differ by most 0.5%. If the allocation rate is over 20%, the IQR from 25 patients reduces from 6-7% to 4-5% for 50 patients. Since the highest allocations were encountered at doses with the greatest persistence, we could conclude an improvement in estimation at these levels.

Lastly, we compare our two-stage adaptive model with the mTPI design, also an adaptive method shown to be safer (fewer patients above the MTD) and with higher probabilities of identifying the MTD as compared to the '3+3' design. Figures 3-6 display the percent allocation per dose for both of these designs, structured by toxicity scenarios.

When all doses have unacceptably high toxicity (Tox1), regardless of the efficacy trend, the mTPI stops after allocating the first cohort of three patients at the lowest dose, whereas the adaptive model allocates approximately 60% at this level. This is expected given the toxicity scenario: with an unacceptable DLT rate of 0.40 and an acceptable rate of 0.15, the

likelihood ratio when one patient in three has a DLT is 0.75, which is not small enough to halt escalation. Note, however, that two DLTs in three patients is considered sufficiently strong evidence of toxicity ($LR = 0.20$) or three DLTs in six patients ($LR = 0.15$) and the dose is deemed too toxic for further evaluation. Also note that if escalation when one patient in three has a DLT is considered too aggressive, the toxicity levels can be altered when designing the trial. For example, if one were to declare the toxic and acceptably toxic levels as 0.30 and 0.08, respectively, then the LR associated with one DLT in three patients is 0.46 and thus the dose would be declared too toxic for further consideration based on $k = 2$.

When the first three doses have low toxicity ($< 15\%$ as in Tox3), the adaptive design assigns most of the patients to the dose with the highest T-cell persistence, i.e., dose 3 for scenarios Eff1-Eff3, and doses 1 and 2 for the plateau trend Eff4 (Figure 5). Not surprisingly, under the same toxicity scenario, the mTPI design is less efficient by allocating most of the patients at dose level 4 (ambiguous toxicity 0.15 - 0.40) before deciding to de-escalate. In general, the mTPI design will allocate most patients to the highest acceptable dose level, regardless of the efficacy profile. This is expected since the mTPI allocation is driven by toxicity only, with the ultimate goal of identifying the MTD. Actually, it has been shown and also confirmed in our simulations that the mTPI has high probability ($>40\%$) of selecting the MTD only in scenarios where at least one of the doses has a toxicity close to the targeted toxicity or the MTD is at the lower or higher end of the dosing set (Table S2). In cancer immunotherapies however, the toxicity profile can be low(er) and the MTD might not necessarily be the most efficacious dose. This issue is being addressed by the adaptive design that incorporates both toxicity and efficacy-based measures and skews the allocation distribution toward acceptable and promising doses. In terms of sample size, the two designs are comparable with a difference of ± 3 patients depending on the toxicity scenario (Table S2).

5 Discussion

In recent years, and especially with the increased number of promising immunotherapies and targeted agents, toxicity profiles are milder at efficacious doses and many of these agents do not adhere to the traditional assumptions including the assumptions that the higher the dose, the greater the efficacy, and that there is a monotonically increasing relationship between dose and efficacy. This has been demonstrated in the phase I trials of pembrolizumab [Patnaik et al. 2015] and nivolumab [Topalian et al. 2012] where there are rather flat dose-toxicity and flat dose-efficacy curves. Recent papers have emphasized the changing paradigm of phase I trials in which both toxicity and efficacy are evaluated, often in multiple disease cohorts [Prowell et al 2016] or called for a movement away from “MTD-based” dose selection with increased interest in dose ranging studies [Ratain et al. 2008]. So, while for chemotherapeutics, toxicity-based dose-finding has made sense, with many of the newer types of therapies, dose-finding should be focused on both toxicity (to ensure doses are not overly toxic) and efficacy-based measures. This is a major shift in oncology drug development and we believe that the proposed design that incorporates dose-finding based on immunotherapeutic response while maintaining acceptable toxicity levels is quite timely.

In this paper we propose an early-phase adaptive design with applications in oncology trials. The dose selection process is based on both safety and immunologic responses, modeled independently, with a monotonicity assumption imposed only on the dose-toxicity relationship. Under these conditions, the study protocol should provide clear guidance on how to distinguish and proceed further in the case of unacceptable toxicity (serious adverse events) or occurrence of toxicity as an indicator of efficacy response. The two-stage process starts with establishing the safety profile of the prespecified doses, followed by an efficacy-driven adaptive randomization. The likelihood approach in the first stage can be easily implemented in any dose-finding trials when the dose is targeted using an efficacy-based outcome. Appropriate for testing simple hypotheses, the likelihood method can accommodate any cohort size per dose, provide consistent decision rules concerning safety at any point in the trial, and notably, use a similar total sample size to the mTPI design.

Stage 1 is also used for collecting immunologic (efficacy) responses. In our application, the immunological outcome is defined as percent persistence of transduced T-cells at follow-up compared to baseline, but the design can easily be implemented to other continuous outcomes. For adoptive T-cell transfer (ACT), data from animal studies indicate a direct correlation between the number of transduced T-cells and antitumor response in vivo [Gattinoni et al. 2005, Wang et al. 2004]. Another ACT study involving humans shows that there is a clear strong correlation between clinical response and persistence of administered cells at one month [Johnson et al. 2009]. Up to this point, this correlation has yet to be translated into prediction of clinical outcome. Still, with several ongoing trials addressing the adoptive cellular approach, we are optimistic that surrogacy is soon to be established given the strong mechanistic hypotheses. Ideally, a strong (surrogate) association should exist between the efficacy outcome and the clinical outcome of interest, which would be the primary endpoint in a phase II or III trial (e.g., progression-free survival, overall response rate, and overall survival).

In our application, the follow-up time for measuring T-cell persistence is 30 days from infusion. Nevertheless, the lag time may vary depending on the area of application. If there is rapid accrual, the randomization probabilities in stage 2 will not update as frequently, and the design will lean more towards balanced allocation across doses. Another important issue regards non-evaluable/nonexistent measurements. These can emerge from drop-outs or other anatomic factors such as impossibility to measure percent persistence because of a total migration of transduced T-cells into the tumor [Rosenberg et al. 1998]. Until now, little effort has been made to understand and solve the problem of missing data in the early phase of drug development. At this point, our model is based on the assumption that all immunological responses are evaluable at follow-up. This is not a limitation of the design since neither stage 1 nor stage 2 requires equal cohort sizes. Still, guidance should be provided in the trial protocol for how to proceed if missing data arise.

The design proposed in this paper performs better than mTPI design, which despite its adaptive nature does not accommodate the new assumptions imposed by targeted or immunotherapy agents. The adaptive design assigns more patients at acceptable doses with higher persistence (efficacy). Despite the relatively small sample size ($N=25$) that inherently yields large variability, these operating characteristics are still promising. Given

the recent trend toward phase I trials with larger sample sizes, the operating characteristics would be further improved [Khoja et al. 2015, Topalian et al. 2012). We demonstrated that by using an efficacy-driven adaptive randomization with safety constraints, the allocation distribution is skewed towards more efficacious doses, and thus limit the number of patients being exposed to dose levels that may potentially be too toxic or have no therapeutic benefit. There are several possible directions for future research. First, we could add additional constraints for doses with similar randomization probabilities, e.g., favor the dose having the smaller number of patients when efficacy values are approximately the same. Secondly, when no DLT is expected or achieved, we could implement other optimality criteria such as: identify doses that pass a minimum efficacy boundary, or are within a certain percentage from a targeted efficacy level. Lastly, rather than using just immunological responses, the dose selection process at the end of the trial could also incorporate a clinical outcome (e.g., progression-free survival). While there is still much to be learned about cancer immunotherapies, there is reason to believe that more will soon be established as standard of care. In this regard, our model provides an ideal setting for dose-finding trials testing cancer immunotherapeutic agents.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This work was partially supported by National Institute of Health (grant number P01 CA154778-01). The authors thank Drs. Sharon D. Yeatts and Viswanathan Ramakrishnan (Department of Public Health Sciences, Medical University of South Carolina) for the constructive comments provided during manuscript production.

References

- [Herberman 1985] Herberman RB Design of clinical trials with biological response modifiers. *Cancer Treatment Reports* 1985, 69:1161–4. [PubMed: 2412692]
- [Roberts et al. 2004] Roberts TG, Goulart BH, Squitieri L, et al. Trends in the risks and benefits to patients with cancer participating in phase I clinical trials. *Journal of the American Medical Association* 2004, 292(17):2130–40. [PubMed: 15523074]
- [Horstmann et al. 2005] Horstmann E, McCabe MS, Grochow L, et al. Risks and benefits of phase I oncology trials, 1991 through 2002. *New England Journal of Medicine* 2005, 352(9):895–904. [PubMed: 15745980]
- [Van Meter, Garrett-Mayer, Bandyopadhyay 2012] Van Meter EM, Garrett-Mayer E, Bandyopadhyay D Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method. *Clinical Trials* 2012; 9(3):303–13. [PubMed: 22547420]
- [Yuan, Chappell, Bailey 2007] Yuan Z, Chappell R, Bailey H The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach. *Biometrics* 2007; 63(1): 173–9. [PubMed: 17447942]
- [Storer 1989] Storer BE. Design and analysis of phase I clinical trials. *Biometrics* 1989; 45(3):925–937. [PubMed: 2790129]
- [O’Quigley, Pepe, Fisher 1990] O’Quigley J, Pepe M, Fisher L Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990, 46(1):33–48. [PubMed: 2350571]

- [Piantadosi, Fisher, Grossman 1998] Piantadosi S, Fisher JD, Grossman S Practical implementation of a modified continual reassessment method for dose-finding trials. *Cancer Chemotherapy and Pharmacology* 1998, 41(6):429–36. [PubMed: 9554585]
- [Moller 1995] Moller S An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses. *Statistics in Medicine* 1995, 14(9 10):911–22. [PubMed: 7569510]
- [Cheung, Chappell 2000] Cheung YK, Chappell R Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000, 56(4): 1177–82. [PubMed: 11129476]
- [Dahlberg et al. 2014] Dahlberg SE, Shapiro GI, Clark JW, Johnson BE Evaluation of statistical designs in phase I expansion cohorts: The Dana-Farber/Harvard Cancer Center experience. *Journal of the National Cancer Institute* 2014; 106(7).
- [Thall, Cook 2004] Thall PF, Cook JD Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 2004, 60(3):684–93. [PubMed: 15339291]
- [Thall, Cook, Estey 2006] Thall PF, Cook JD, Estey EH Adaptive dose selection using efficacy-toxicity trade-offs: illustrations and practical considerations. *Journal of Biopharmaceutical Statistics* 2006, 16(5):623–38. [PubMed: 17037262]
- [Thall, Russell 1998] Thall PF, Russell KE A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* 1998, 54(1):251–64. [PubMed: 9544520]
- [Braun 2002] Braun TM. The bivariate continual reassessment method. Extending the CRM to phase I trials of two competing outcomes. *Controlled Clinical Trials* 2002, 23(3):240–56. [PubMed: 12057877]
- [Zhang, Sargent, Mandrekar 2006] Zhang W, Sargent DJ, Mandrekar S An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine* 2006, 25(14):2365–83. [PubMed: 16220478]
- [Hoering, LeBlanc, Crowley 2011] Hoering A, LeBlanc M, Crowley J Seamless phase I-II trial design for assessing toxicity and efficacy or targeted agents. *Clinical Cancer Research* 2011, 17(4):640–6. [PubMed: 21135145]
- [Messer et al. 2010] Messer K, Natarajan L, Ball ED, et al. Toxicity-evaluation designs for phase I/II cancer immunotherapy trials. *Statistics in Medicine* 2010, 29(7–8):712–20. [PubMed: 20213706]
- [Ji and Wang 2013] Ji Y, Wang SJ. Modified toxicity probability interval design: a safer and more reliable method than the 3+3 design for practical phase I trials. *Journal of Clinical Oncology* 2013, 31:1785–1791. [PubMed: 23569307]
- [Hacking 1965] Hacking I *Logic of statistical inference*. London: Cambridge University Press, 1965.
- [Royall 1997] Royall R *Statistical Evidence: A likelihood paradigm*. London: Chapman Hall/CRC, 1997.
- [Blume 2002] Blume JD. Likelihood methods for measuring statistical evidence. *Statistics in Medicine* 2002; 21(17):2563–2599. [PubMed: 12205699]
- [Royall 2000] Royall R On the probability of observing misleading statistical evidence. *Journal of the American Medical Association* 2000; 95(451):760–768.
- [Chiuzan, Garrett-Mayer, Yeatts 2015] Chiuzan C, Garrett-Mayer E, Yeatts S A likelihood-based approach for computing the operating characteristics of the 3+3 phase I clinical trial design with extensions to other A+B designs. *Clinical Trials* 2015; 12(1):24–33. [PubMed: 25349178]
- [Patnaik et al. 2015] Patnaik A, Kang SP, Rasco D, Papadopoulos KP, Elassaiss-Schaap J, et al. Phase I study of pembrolizumab (MK-3475; anti-PD-1 monoclonal antibody) in patients with advanced solid tumors. *Clinical Cancer Research* 2015, 21(19):4286–93. [PubMed: 25977344]
- [Topalian et al. 2012] Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *New England Journal of Medicine* 2012;366: 2443–54. [PubMed: 22658127]
- [Prowell et al. 2016] Prowell TM, Theoret MR and Pazdur R. Seamless oncology-drug development. *New England Journal of Medicine* 2016; 374:2001–2003. [PubMed: 27074059]
- [Ratain et al. 2008] Ratain MJ, Humphrey RW, Gordon GB et al. Recommended Changes to Oncology Clinical Trial Design: Revolution or Evolution? *European Journal of Cancer* 2008; 44(1): 8–11. [PubMed: 17981025]

- [Gattinoni et al. 2005] Gattinoni L, Klebanoff CA, Palmer DC, et al. Acquisition of full effector function in vitro paradoxically impairs the in vivo antitumor efficacy of adoptively transferred CD8+ T cells. *Journal of Clinical Investigation* 2005; 115:1616–1626. [PubMed: 15931392]
- [Wang et al. 2004] Wang LX, Huang LX, Graor H, et al. Adoptive immunotherapy of cancer with polyclonal, 108-fold hyperexpanded, CD4+ and CD8+ T cells. *Journal of Translational Medicine* 2004; 2:41. [PubMed: 15566571]
- [Johnson et al. 2009] Johnson LA, Morgan RA, Dudley ME, et al. Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen. *Blood* 2009; 114(3):535–546. [PubMed: 19451549]
- [Rosenberg et al. 1998] Rosenberg SA, Yang JC, Schwartzentruber DJ, et al. Immunologic and therapeutic evaluation of a synthetic peptide vaccine for the treatment of patients with metastatic melanoma. *Nature Medicine* 1998; 4:321–327.
- [Khoja et al. 2015] Khoja L, Butler MO, Kang SP, et al. Pembrolizumab. *J Immunother Cancer* 2015; 3:36. [PubMed: 26288737]

Figure 1: Illustration of variability among doses using a beta-distribution variance of 0.2% (A) versus 1% (B). A plateau trend for T-cell persistence as function of dose (0.05, 0.25, 0.65, 0.65, 0.65) and a total sample size of 25 patients are used in both cases.

Figure 2: Simulated efficacy (A) and toxicity (B) scenarios as a function of dose level. The efficacy outcome is defined as persistence (%) of transduced T-cells at follow-up compared to baseline.

Figure 3:

Percent allocation per dose for the two-stage adaptive design versus mTPI design. Results of 5,000 simulated trials are summarized as: median (filled circle/square), 25th-75th percentiles (solid line), and 2.5th-97.5th percentiles (dashed line). Eff1=(0.05, 0.15, 0.40, 0.65, 0.80), Eff2=(0.05, 0.25, 0.70, 0.65, 0.20), Eff3=(0.05, 0.25, 0.65, 0.65, 0.65), Eff4=(0.60, 0.60, 0.61, 0.62, 0.60).

Figure 4:

Percent allocation per dose for the two-stage adaptive design versus mTPI design. Results of 5,000 simulated trials are summarized as: median (filled circle/square), 25th-75th percentiles (solid line), and 2.5th-97.5th percentiles (dashed line). Eff1=(0.05, 0.15, 0.40, 0.65, 0.80), Eff2=(0.05, 0.25, 0.70, 0.65, 0.20), Eff3=(0.05, 0.25, 0.65, 0.65, 0.65), Eff4=(0.60, 0.60, 0.61, 0.62, 0.60)

Figure 5:

Percent allocation per dose for the two-stage adaptive design versus mTPI design. Results of 5,000 simulated trials are summarized as: median (filled circle/square), 25th-75th percentiles (solid line), and 2.5th-97.5th percentiles (dashed line). Eff1=(0.05, 0.15, 0.40, 0.65, 0.80), Eff2=(0.05, 0.25, 0.70, 0.65, 0.20), Eff3=(0.05, 0.25, 0.65, 0.65, 0.65), Eff4=(0.60, 0.60, 0.61, 0.62, 0.60).

Figure 6:

Percent allocation per dose for the two-stage adaptive design versus mTPI design. Results of 5,000 simulated trials are summarized as: median (filled circle/square), 25th-75th percentiles (solid line), and 2.5th-97.5th percentiles (dashed line). Eff1=(0.05, 0.15, 0.40, 0.65, 0.80), Eff2=(0.05, 0.25, 0.70, 0.65, 0.20), Eff3=(0.05, 0.25, 0.65, 0.65, 0.65), Eff4=(0.60, 0.60, 0.61, 0.62, 0.60).

Table 1:

Percent allocation (%) per dose at the end of stage 2. Results are summarized as median (interquartile range).

Toxicity [†]	Efficacy [‡]	N [§]	D ₁	D ₂	D ₃	D ₄	D ₅
Tox1	Eff1	25 50	56 (24-100) 53 (29-100)	20(0-25) 23(0-25)	12 (0-20) 10 (0-19)	0 (0-15) 0 (0-13)	0 (1-12) 0 (0-7)
	Eff2	25 50	56 (24-100) 57 (30-100)	22 (0-25) 25 (0-27)	12 (0-20) 9 (0-19)	0 (0-13) 0 (0-12)	0 (0-12) 0 (0-6)
	Eff3	25 50	56 (24-100) 57 (29-100)	24 (0-25) 23 (0-26)	12 (0-20) 9 (0-19)	0 (0-14) 0 (0-13)	0 (0-12) 0 (0-7)
	Eff4	25 50	56 (32-100) 56 (35-100)	21 (0-25) 22 (0-25)	12 (0-19) 9 (0-19)	0 (0-15) 0 (0-14)	0 (0-12) 0 (0-8)
Tox2	Eff1	25 50	20 (12-32) 22 (14-32)	25 (17-36) 32 (22-56)	20 (15-32) 16 (8-32)	12 (0-20) 8 (0-13)	0 (0-12) 0 (0-7)
	Eff2	25 50	16 (12-28) 16 (12-31)	25 (20-40) 36 (24-62)	20 (14-32) 16 (8-32)	12 (0-16) 8 (0-13)	0 (0-12) 0 (0-6)
	Eff3	25 50	16 (12-28) 16 (12-30)	25 (20-40) 36 (24-62)	20 (14-32) 16 (8-32)	12 (0-16) 8 (0-13)	0 (0-12) 0 (0-6)
	Eff4	25 50	32 (24-43) 38 (30-46)	25 (30-32) 30 (25-40)	18 (13-24) 16 (8-23)	12 (0-16) 7 (0-13)	0 (0-12) 0 (0-6)
Tox3	Eff1	25 50	16 (12-16) 10 (8-14)	16 (12-24) 18 (14-26)	24 (20-36) 34 (22-50)	20 (16-28) 14 (8-32)	12 (12-20) 6 (6-10)
	Eff2	25 50	12 (12-16) 10 (8-12)	20 (16-24) 20 (16-26)	29 (24-36) 40 (30-52)	20 (16-28) 14 (8-28)	12 (12-16) 6 (6-8)
	Eff3	25 50	12 (12-16) 10 (8-12)	20 (16-24) 20 (16-28)	28 (20-36) 38 (28-50)	20 (12-24) 14 (8-26)	12 (12-16) 6 (6-10)
	Eff4	25 50	24 (20-28) 26 (22-32)	24 (20-28) 26 (22-32)	20 (16-28) 24 (18-28)	16 (12-20) 14 (8-20)	12 (12-16) 6 (6-10)
Tox4	Eff1	25 50	12 (12-16) 10 (8-12)	12 (12-16) 10 (8-12)	20 (16-24) 20 (16-22)	24 (20-28) 28 (24-32)	28 (24-32) 32 (28-36)
	Eff2	25 50	12 (12-16) 10 (8-12)	16 (12-20) 14 (12-18)	28 (24-32) 32 (28-36)	24 (20-28) 30 (26-34)	16 (12-20) 12 (10-16)
	Eff3	25 50	12 (12-16) 10 (8-12)	16 (12-20) 13 (10-16)	24 (20-28) 26 (22-30)	24 (20-28) 26 (22-30)	24 (20-28) 26 (22-30)
	Eff4	25 50	20 (16-24) 20 (16-24)	20 (16-24) 20 (16-24)	20 (16-24) 20 (16-24)	20 (16-24) 20 (16-24)	20 (16-24) 20 (16-24)

[†]Tox1=(0.35, 0.35, 0.35, 0.35, 0.35), Tox2=(0.05, 0.18, 0.29, 0.40, 0.51), Tox3=(0.02, 0.05, 0.12, 0.27, 0.40), Tox4=(0.02, 0.02, 0.02, 0.02, 0.02)

[‡]Eff1=(0.05, 0.25, 0.70, 0.65, 0.20), Eff2=(0.05, 0.25, 0.70, 0.65, 0.20), Eff3=(0.05, 0.25, 0.65, 0.65, 0.65) Eff4=(0.60, 0.60, 0.61, 0.62, 0.60)

[§]N represents the maximum sample size (stage 1 + stage 2) used in simulations.

Table 2:

Estimated percent efficacy (%) per dose at the end of stage 2 of the adaptive model. Results are summarized as median (interquartile range).

Toxicity [†]	N [§]	Eff1 (true mean efficacy) [‡]				
		D ₁ (5)	D ₂ (15)	D ₃ (40)	D ₄ (65)	D ₅ (80)
Tox1	25	2.0 (1.0-5.0)	13.0 (9.5-17.0)	40.0 (36.0-44.0)	66.0 (61.0-70.0)	81.0 (77.0-85.0)
	50	2.0 (1.0-5.0)	13.0 (10.0-17.0)	39.5 (36.0-43.5)	65.0 (61.0-70.0)	81.0 (77.0-85.0)
Tox2	25	2.0 (1.0-4.5)	13.0 (10.0-16.0)	40.0 (36.0-43.5)	65.0 (61.0-70.0)	81.0 (76.5-85.0)
	50	2.0 (1.0-4.0)	13.0 (10.5-15.0)	40.0 (36.5-43.0)	65.5 (61.0-70.0)	81.0 (77.0-85.0)
Tox3	25	2.0 (1.0-5.0)	13.0 (9.5-17.0)	39.5 (36.0-43.0)	65.5 (61.0-69.0)	81.5 (77.0-85.0)
	50	2.0 (1.0-4.0)	13.0 (10.5-15.5)	39.5 (37.0-42.0)	65.5 (62.0-69.0)	81.5 (77.0-85.0)
Tox4	25	2.5 (1.0-5.0)	13.0 (9.0-17.5)	40.0 (35.5-44.0)	65.0 (62.0-69.0)	81.5 (78.0-85.0)
	50	2.5 (1.0-4.5)	13.0 (9.5-16.5)	40.0 (37.0-43.0)	65.5 (63.0-68.0)	81.5 (79.0-84.0)
		Eff2 (true mean efficacy)				
		D ₁ (5)	D ₂ (25)	D ₃ (70)	D ₄ (65)	D ₅ (20)
Tox1	25	2.5 (1.0-5.0)	24.0 (20.0-28.0)	71.0 (67.0-74.5)	65.5 (61.0-70.0)	19.0 (14.4-23.0)
	50	2.0 (1.0-5.0)	24.0 (20.5-27.5)	71.0 (67.0-74.0)	66.0 (61.0-69.5)	19.0 (14.0-23.5)
Tox2	25	2.5 (1.0-5.0)	24.0 (21.0-27.5)	70.5 (67.0-74.0)	66.0 (61.0-70.0)	18.3 (14.0-23.0)
	50	2.0 (1.0-4.0)	24.0 (22.0-26.0)	71.0 (67.0-74.0)	65.0 (61.0-70.0)	18.0 (14.0-23.5)
Tox3	25	2.5 (1.0-5.0)	24.0 (20.0-28.0)	71.0 (67.0-74.0)	65.5 (61.1-69.0)	18.5 (14.0-23.0)
	50	2.5 (1.0-4.5)	24.0 (21.0-26.5)	71.0 (68.5-73.0)	65.5 (62.0-69.0)	19.0 (14.0-23.0)
Tox4	25	2.5 (1.0-5.0)	24.0 (20.0-28.0)	70.5 (67.0-74.0)	65.5 (62.0-69.0)	18.5 (14.5-23.0)
	50	2.5 (1.0-4.5)	24.0 (20.5-27.0)	71.0 (68.0-73.0)	65.0 (63.0-68.0)	18.5 (15.0-22.0)
		Eff1 (true mean efficacy)				
		D ₁ (5)	D ₂ (25)	D ₃ (65)	D ₄ (65)	D ₅ (65)
Tox1	25	2.5 (1.0-5.0)	24.0 (20.0-28.0)	66.0 (61.0-70.0)	65.0 (61.0-70.0)	66.0 (61.0-70.0)
	50	2.0 (1.0-6.0)	24.0 (20.0-27.5)	65.5 (61.5-69.0)	65.5 (61.1-69.0)	66.0 (62.0-70.0)
Tox2	25	2.0 (1.0-5.0)	24.0 (20.5-27.0)	65.5 (62.0-69.0)	65.5 (61.0-69.5)	65.5 (61.0-70.0)
	50	2.0 (1.0-4.0)	24.0 (22.0-26.0)	65.5 (62.0-69.5)	65.0 (61.0-69.0)	66.0 (61.0-70.0)
Tox3	25	2.0 (1.0-5.0)	24.0 (20.0-28.0)	65.5 (62.0-69.0)	65.5 (61.0-70.0)	65.0 (61.0-70.0)
	50	2.5 (1.0-4.0)	24.0 (21.0-27.0)	65.5 (63.0-68.0)	65.5 (62.0-69.0)	65.5 (61.0-69.5)
Tox4	25	2.0 (1.0-5.0)	24.0 (20.0-28.0)	65.0 (61.5-69.0)	65.5 (61.5-69.0)	65.5 (61.5-69.0)
	50	2.5 (1.0-4.5)	24.0 (20.0-27.0)	65.5 (63.0-68.0)	65.0 (63.0-68.0)	65.5 (63.0-68.0)
		Eff4 (true mean efficacy)				
		D ₁ (60)	D ₂ (60)	D ₃ (61)	D ₄ (62)	D ₅ (60)
Tox1	25	60.5 (55.5-65.0)	60.0 (55.5-64.0)	61.0 (57.0-66.0)	62.0 (58.0-66.5)	60.0 (56.0-65.0)
	50	60.0 (56.0-65.0)	60.0 (56.5-64.5)	61.0 (57.0-65.0)	62.0 (58.0-67.0)	60.0 (55.0-65.0)
Tox2	25	60.0 (56.5-64.0)	60.0 (56.5-64.0)	61.5 (57.0-65.0)	62.0 (58.0-67.0)	60.0 (56.0-65.0)
	50	60.0 (57.5-63.0)	60.0 (57.5-63.0)	61.5 (58.0-65.0)	62.0 (58.0-66.5)	60.0 (55.0-65.0)
Tox3	25	60.0 (56.0-64.0)	60.0 (56.0-64.0)	61.5 (57.0-65.0)	62.0 (58.0-66.0)	60.0 (56.0-65.0)
	50	60.0 (57.5-63.0)	60.0 (57.5-63.0)	61.5 (58.0-64.0)	62.0 (59.0-66.)	60.0 (56.0-64.0)
Tox4	25	60.0 (56.5-64.0)	60.0 (56.5-64.0)	61.0 (57.0-65.0)	62.0 (58.0-66.0)	60.0 (56.0-64.0)
	50	60.0 (57.0-63.0)	60.0 (57.0-63.0)	61.0 (58.5-6540)	62.5 (59.5-65.0)	60.0 (57.0-63.0)

[†]Tox1=(0.35, 0.35, 0.35, 0.35, 0.35), Tox2=(0.05, 0.18, 0.29, 0.40, 0.51), Tox3=(0.02, 0.05, 0.12, 0.27, 0.40), Tox4= (0.02, 0.02, 0.02, 0.02, 0.02)

[§]N represents the maximum sample size (stage 1 + stage 2) used in simulations.

[‡]Efficacy is defined as percent persistence of transduced T-cells at follow-up compared to baseline. Eff1-Eff4 show the mean persistence values (for each dose) used in simulations.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript