Memorial Sloan-Kettering Cancer Center
Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology
& Biostatistics Working Paper Series

Year 2014 *Paper* 28

Dose expansion cohorts in Phase I trials

Alexia Iasonos* John O'Quigley†

*Memorial Sloan Kettering Cancer Center, iasonosa@mskcc.org

†Universit'e Pierre et Marie Curie-Paris VI, Paris, France, john.oquigley@upmc.fr

http://biostats.bepress.com/mskccbiostat/paper28

Dose expansion cohorts in Phase I trials

Alexia Iasonos and John O'Quigley

Abstract

A rapidly increasing number of Phase I dose-finding studies, and in particular those based on the standard 3+3 design, frequently prolong the study and include dose expansion cohorts (DEC) with the goal to better characterize the toxicity profiles of experimental agents and to study disease specific cohorts. These trials consist of two phases: the usual dose escalation phase that aims to establish the maximum tolerated dose (MTD) and the dose expansion phase that accrues additional patients, often with different eligibility criteria, and where additional information is being collected. Current protocols typically do not specify whether the MTD will be updated in light of the new data accumulated from the DEC. In this paper, we propose methodology that allows monitoring of safety in the DEC by re-evaluating the MTD in light of additional information. Our working assumption is that, regardless of the design being used for dose escalation, during the DEC we are experimenting in the neighborhood of a target dose with an acceptable rate of toxicity. We refine our initial estimate of the MTD by continuing experimentation in the immediate vicinity of the initial estimate of the MTD. The auxiliary information provided in this evaluation can include toxicity, pharmacokinetic, efficacy or other endpoints. Weconsider approaches specifically focused on the aims of DEC, that examine efficacy alone or simultaneously with safety and compare the proposed tests via simulations.

Dose expansion cohorts in Phase I trials

Alexia Iasonos[1] and John O'Quigley[2]

(1) Department of Biostatistics, Memorial Sloan Kettering Cancer Center, New York, USA.

(2) Université Pierre et Marie Curie-Paris VI, Paris, France.

Abstract

A rapidly increasing number of Phase I dose-finding studies, and in particular those based on the standard 3+3 design, frequently prolong the study and include dose expansion cohorts (DEC) with the goal to better characterize the toxicity profiles of experimental agents and to study disease specific cohorts. These trials consist of two phases: the usual dose escalation phase that aims to establish the maximum tolerated dose (MTD) and the dose expansion phase that accrues additional patients, often with different eligibility criteria, and where additional information is being collected. Current protocols typically do not specify whether the MTD will be updated in light of the new data accumulated from the DEC. In this paper, we propose methodology that allows monitoring of safety in the DEC by re-evaluating the MTD in light of additional information. Our working assumption is that, regardless of the design being used for dose escalation, during the DEC we are experimenting in the neighborhood of a target dose with an acceptable rate of toxicity. We refine our initial estimate of the MTD by continuing experimentation in the immediate vicinity of the initial estimate of the MTD. The auxiliary information provided in this evaluation can include toxicity, pharmacokinetic, efficacy or other endpoints. We consider approaches specifically focused on the aims of DEC, that examine efficacy alone or simultaneously with safety and compare the proposed tests via simulations. KEYWORDS: dose finding; Phase I trials; dose expansion; sequential monitoring; average sample number.

1 Introduction

Phase I trials are increasingly using dose expansion cohorts to better characterize the toxicity profiles of experimental agents or to study disease specific cohorts before selecting an appropriate dose and patient population upon which closer attention will be focused for the Phase II study [Iasonos and O'Quigley 2013, Manji et al. 2013]. The experimental setting for the dose expansion cohort (DEC) differs in two important ways from that of the Phase I study. The first difference is the recruitment criteria. The DEC patients are likely to belong to a more narrowly and more sharply defined targeted category of patients, for example disease specific or histology specific cohorts. The second difference relates to the information gathered on the DEC patients. This is typically broader than that for the Phase I and usually will include efficacy data as well as additional information concerning toxicity gathered during the Phase I stage. Furthermore, Phase I trials with dose expansion cohorts do not have the same objectives as Phase I/II trials so that Phase I trials with DEC raise numerous, new design considerations as they fall somewhere in between Phase I, Phase I/II or Phase II trials (Figure 1). There is currently no design specific to DEC [Manji et al. 2013, Iasonos and O'Quigley 2013], a shortcoming recently identified by a National Institute of Health working group.

In this paper, we propose methodology that takes into account the information provided from the additional patients treated at the maximum tolerated dose (MTD) as part of an expansion cohort and we re-evaluate the recommended Phase II dose (RP2D) based on the information provided by all the data. The proposed design will provide efficacy estimates on multiple levels, so that investigators can

decide which dose to take forward to a phase II testing. Over the last 20 years much work has been done on model based designs for Phase I dose-finding studies [O'Quigley, Pepe and Fisher 1990, Iasonos et al. 2008, Yuan et al 2011], which includes methods that can simultaneously deal with a bivariate outcome of toxicity and efficacy [Yuan et al 2011, O'Quigley, Hughes and Fenton 2001]. Phase I dose-finding studies carried out according to a model-based design are very easily adapted to take on board the additional dose-expansion cohort. Whenever possible we would recommend that any Phase I dose-finding trial uses one of the model based approaches, for several reasons including the ability to readily extend to situations involving an expansion cohort. However, the algorithmic 3+3 design, despite its known poor properties still heavily dominates the field and continues to be used in approximately 90% of new studies [Rogatko et al. 2007]. At Memorial Sloan Kettering Cancer Center alone, there are approximately 130 such trials annually that make use of the 3+3 design. Despite clinical investigators' reluctance to move away from the 3+3 design, analysis methods for the DEC are needed since the 3+3 no longer applies. Current practice consists in simply treating patients at the estimated MTD with no planned analysis of the observed outcomes of the patients accrued during DEC. Our purpose is to develop designs for DEC that will be applicable regardless of the type of design used in the Phase I stage. This includes Phase I trials based on the 3+3 which account for the great majority of current trials.

Phase I protocols with dose expansions consist of two phases: the dose escalation phase followed by a dose expansion phase (Figure 1). During the dose escalation phase, the MTD is established and during the expansion phase, an additional number of patients (6, 10 or more) are treated at the estimated MTD [Topalian et al. 2012].

While the MTD is defined by the dose escalation phase, the RP2D is based on the combined safety data (pre and post dose expansion) and it is the dose level selected for future trials. The toxicity observed from the additional patients being treated at the expansion phase, while being reviewed by clinical investigators, is not part of the dose escalation algorithm that was followed to establish the MTD. There are cases where the RP2D is lower or higher than the MTD [Manji et al. 2013, Isambert et al. 2012] given the safety evaluation during the expansion phase, pharmacokinetic studies or efficacy endpoints. The basis of our paper is that there is still considerable uncertainty in the selection of the MTD after the dose escalation part, and zooming into the vicinity of the MTD while exploring other endpoints beyond safety will help guide the choice of the RP2D.

The aims of Phase I trials with DEC can be twofold depending on whether the objective is to recommend the MTD based on safety alone, or whether the aim is to further investigate for evidence of efficacy in a selected patient population. Combining safety and efficacy assessments in clinical trials in oncology under the heading of a single study or protocol has been relatively common and statistical designs for Phase I/II trials exist [Yuan et al 2011, O'Quigley, Hughes and Fenton 2001]. Unlike Phase I/II trials where the Phase II part of the protocol aims to target a dose with a specific efficacy threshold, Phase I trials with DEC are considered more exploratory in nature and often no efficacy is being measured during pre-expansion. As such, Phase I protocols with expansion cohorts often aim to establish a safe dose and at the same time obtain initial evidence whether to take the investigational drug forward to a larger Phase II study and if so, at which dose level. Thus, DEC do not necessarily need to include all patients at a single dose level, or target a dose with an

efficacy threshold as in the Phase II setting, but instead, can further explore more than a single dose level. At the end of the study, after accruing a number of additional patients at the expansion phase there might be an indication that the MTD chosen based on safety alone may likely have low efficacy. The recommendation then might be that investigators either need to add additional patients at the expansion phase to better estimate the efficacy rate or that a higher dose should be considered which is more likely to show activity. Alternatively, another approach will be to test multiple levels in the expansion phase with the aim of selecting the dose with a higher efficacy rate. Whether the efficacious level is unsafe and abandonment of the investigational drug should be considered altogether is one of the questions the proposed methodology aims to address. In the following sections, we develop these ideas in a more formal setting.

2 Methods

The main idea is to use some model based design and guide the dose allocation of the expansion cohort based on the complete data that consists of the data prior and post expansion. Our focus here is the Continual Reassessment Method (CRM) [O'Quigley, Pepe and Fisher 1990], although we may employ any other model based design. CRM assumes the dose-toxicity curve can be modeled through a simple working model and as accumulated data on patients' responses are obtained, we update this model, and assign the next patient at a level closest to an acceptable toxicity rate, θ, based on some measure of distance. We assume the trial consists of k ordered dose levels, $d_1, d_2, \ldots, d_k$, and a total of N patients. The assigned dose

level for patient j is denoted as X_j, and the binary toxicity outcome is denoted as Y_j, where Y_j= 1 indicates a DLT for patient j, and 0 indicates absence of a DLT. We denote the true probability of toxicity at $X_j = x_j$ by: $R(x_j) = \Pr(Y_j = 1|X_j = x_j), \quad x_j \in \{d_1, d_2, \ldots, d_k\}$. O'Quigley and Shen (1996) used a simple working model for the dose toxicity relation of the form, $\psi(d_i, a) = \alpha_i^a$, where $a \in (0, \infty)$ is the unknown parameter, and α_i are the standardized units representing the discrete dose levels d_i or skeleton. Optimal model skeletons can be selected by following the work from Lee and Cheung 2011. Since drugs are assumed to be more toxic at higher dose levels, $\psi(d_i, a)$ is assumed to be an increasing function of d_i. The derivative of log likelihood expressed in terms of dose levels d_i after j patients have been treated, can be expressed as:

$$\mathcal{U}_j(a) \quad = \quad \sum_{i=1}^{k} \left[t_i(j) \frac{\psi'}{\psi}(d_i, a) + (n_i(j) - t_i(j)) \frac{-\psi'}{1-\psi}(d_i, a) \right] \tag{1}$$

where $n_i(j), t_i(j)$ are the number of patients treated and the number of DLTs respectively at each dose level i out of a total of j patients. Once the current estimate of $\hat{a}$ by solving $\mathcal{U}_j(a) = 0$ is obtained, and $\hat{R}(d_i) = \psi(d_i, \hat{a})$ are calculated, the MTD is defined to be the dose $d_m \in \{d_1, \ldots, d_k\}, 1 \leq m \leq k$ such that, $d_m = \arg\min_{d_i} \Delta(\hat{R}(d_i), \theta), \quad i = 1, ..., k.$ where $\Delta(\hat{R}(d_i), \theta)$ denotes the distance from the target acceptable rate θ. For example, in the field of dose finding studies, it is common to use the Euclidean distance, $\Delta(\hat{R}(d_i), \theta) = |\hat{R}(d_i) - \theta|$. Note that m is a random quantity that depends on $\hat{R}(d_i)$ which are random. However conditional on the data $\mathcal{F} = \{(x_j, y_j), j = 1, ...N\}$ from N patients, then d_m is determined at each step.

2.1 Prospective expansion guided by toxicity and/or efficacy

Assume that for the patients accrued during the expansion phase,i.e., in a subset of patients, in addition to monitoring their toxicity we also collect some other measures of efficacy or pharmacodynamic, pharmacokinetic endpoints [Pei and Hughes 2008, O'Quigley, Hughes, Fenton, Pei 2010]. There are studies for example that modify eligibility criteria to require measurable disease or biopsy for patients accrued in the expansion cohort and therefore efficacy is measured on a subset of patients. Accordingly, we re-write the log likelihood function and express it as the sum of two components, the contributions from patients who have toxicity outcomes, this includes all patients, and the contributions of patients who have efficacy measures, whom we assume are the patients accrued during the expansion phase only. Let us denote the true probability of efficacy response at $X_j = x_j$ as: $Q(x_j) = \Pr(V_j = 1|X_j = x_j)$, where V_j is a binary random variable denoting efficacy response for patient j. Similarly as before, we will use a one-parameter working model $\phi(d_i, b) = \beta_i^b$ for $Q(x_j)$ [O'Quigley, Hughes and Fenton 2001], where β_i is a skeleton of initial probabilities of efficacy. We require that for any dose level $d_i \in \{d_1, \ldots, d_k\}$ there exists a value of b , such that $\phi(d_i, b_i) = Q(d_i), i = 1, ..., k$. We assume that N patients have safety measurements, and $J, J \leq N$ patients accrued in the expansion have efficacy and safety measurements. Clearly the toxicity and efficacy outcomes are not independent. One way to proceed would be via the use of copula models in conjunction with standard marginal models for the rates of toxicity and efficacy. Dependence on the particular, often arbitrarily chosen, correlation structure for the copula model can be problematic and so our suggestion is to bypass this by making a simple working assumption that the correlation that exists between toxicity and efficacy is essentially captured

by the given dose. As a consequence, given the dose, these probabilities can now be treated as conditionally independent. We use that as a working assumption, and, in addition, we carry out extensive simulations to investigate the robustness of the inference to significant departures from that assumption. The derivatives of the log likelihood with respect to a and b are given by:

$$\begin{aligned} \partial \log L(a,b|\mathcal{F})/\partial a &= \sum_{i=1}^{k} [t_i(N)\frac{\psi'}{\psi}(d_i,a) + (n_i(N) - t_i(N))\,\frac{-\psi'}{1-\psi}(d_i,a)] \\ \partial \log L(a,b|\mathcal{F})/\partial b &= \sum_{i=1}^{k} [r_i(J)\frac{\phi'}{\phi}(d_i,b) + (n_{e,i}(J) - r_i(J))\,\frac{-\phi'}{1-\phi}(d_i,b)] \end{aligned} \quad (2)$$

where $n_{e,i}(J)$ are the number of patients treated at dose i who are also evaluable for response and $r_i(J)$ are the number of responders observed at dose i.

Instead of assigning the next patient systematically at a dose $x_{j+1} = d_m$, we randomize patients to two levels, d_m and the level just above d_m if $\hat{R}(d_m) < \theta$ or the level just below d_m if $\hat{R}(d_m) > \theta$. We can base the randomization on a random mechanism by using equal randomization probabilities of 0.50 at the two levels or by using the inverse of the distance to θ. In other words, we randomly select a dose according to a discrete distribution which depends on the available levels and the current estimate of the MTD. On the basis of the data from the first j patients, we assign the $(j+1)^{th}$ patient to a dose level as follows:

- If $\hat{R}(d_m) \leq \theta < \hat{R}(d_{m+1}), 1 \leq m < k$ then we randomize the patient to one of two dose levels, $x_{j+1} = d_m$ with probability p_m where

$$p_m = 1 - \left\{ \Delta(\hat{R}\,(d_m)\,,\theta) \left[\Delta(\hat{R}\,(d_{m+1})\,,\theta) + \Delta(\hat{R}\,(d_m)\,,\theta) \right]^{-1} \right\} \quad (3)$$

or $x_{j+1} = d_{m+1}$ with probability $p_{m+1} = 1 - p_m$.

- If $\hat{R}(d_k) < \theta$, we allocate the patient to the last two levels with randomization probabilities of $p_k = p_{k-1} = 0.5$.

- If $\hat{R}(d_1) > \theta$, allocation is still probabilistic but we ensure the closest level, which is the lowest dose, is chosen with higher probability for safety reasons. For example we can use the following probabilities for the lowest two levels: $p_1 = 0.8$ and $p_2 = 0.2$, or by design, as long as $\hat{R}(d_1) > \theta$, we could use $p_1 = 1$.

We continue this algorithm until a fixed number of patients have been accrued during the expansion phase and at the end of the trial the RP2D is defined as the dose closest to the target rate, as in the definition given for d_m, for the $(N+1)^{th}$ dose assignment. Randomizing to two levels, say d_m, d_{m+1}, and assuming the model $\psi(d_m, a)$, then the estimate $\hat{a}$ will converge almost surely to the value a_0 where $U(a_0) = 0$ and $U(a)$ is given by:

$$\begin{aligned} U(a) &= \pi(d_m)[R_m \frac{\psi'}{\psi}(d_m, a) + (1 - R_m)\frac{-\psi'}{1-\psi}(d_m, a)] \\ &+ [1 - \pi(d_m)][R_{m+1}\frac{\psi'}{\psi}(d_{m+1}, a) + (1 - R_{m+1})\frac{-\psi'}{1-\psi}(d_{m+1}, a)] \end{aligned}$$

where $\pi(d_m)$ is the stable distribution of patients included at level d_m [O'Quigley 2006].

The goal of Phase I trials remains to find a safe and acceptable dose and that is the rationale of following the safety criterion along with randomization for dose allocation. Thus far the RP2D is the dose that is chosen with respect to the probability distribution with p_m (Equation 3). The collection of efficacy data is considered a secondary and often exploratory aim as in obtaining some initial information on

efficacy to design further studies. In the next section we give secondary criteria for the definition of the RP2D. The information at the end of the study might not be enough to distinguish levels based on their efficacy rates, but the idea is that these predicted efficacy rates at each level separately based on the above estimation, will provide additional insight for deciding which is the most promising dose to move forward to a Phase II study.

2.2 Monitoring the expansion cohort

Here we formalize the choice of the RP2D based on hypothesis testing for efficacy. The question we address is whether we can reach a definitive conclusion on whether the chosen dose should move forward to later testing given that it meets some promising threshold for efficacy. The efficacy assessment contributes to deciding which dose to take forward as the RP2D as follows. Denote a low efficacy rate, say q_0, and a higher, more clinically interesting rate, q_1, $0 < q_0 < q_1 < 1$. We test the hypotheses: $H_0 : Q(d_i) \leq q_0$ against $H_1 : Q(d_i) \geq q_1$, where $Q(d_i)$ denotes the true, unknown, efficacy rate at level d_i. At each dose level d_i, these hypotheses correspond to $H_0 : b \geq b_0$ against $H_1 : b \leq b_1$ where b is the parameter that models the dose - efficacy relationship. Since these are composite hypotheses, under H_0 we define the region B_0 to be (b_0, ∞) and similarly under H_1 the region B_1 is $(0, b_1)$. The toxicity contributions involving the parameter a cancel out, when the respective integrals cover the regions B_1 and B_0 under the H_1 and H_0 respectively. Specifically, assuming j^* patients have

been treated at level d_i, the test at level d_i is given by:

$$T_1(d_i) = \frac{\int_{B_1} \prod_{l=1}^{j^*} \beta_i^{bv_l}(1-\beta_i^b)^{(1-v_l)} g(b)\, db}{\int_{B_0} \prod_{l=1}^{j^*} \beta_i^{bv_l}(1-\beta_i^b)^{(1-v_l)} g(b)\, db} \tag{4}$$

where $g(b)$ is a pre-specified probability distribution for the parameter b. Note, that the proposed test is calculated at each dose level separately by including contributions to the likelihood from patients treated at the respective level alone. This test will support H_1 if there is sufficient evidence in support of H_1, ie $T_1(d_i) > (1-\epsilon_2)/\epsilon_1$, where the boundaries depend on the choice of Type I and II error rates denoted as ϵ_1, ϵ_2 respectively, and on the sequence of efficacy responses. At the end, we might still be indecisive if the test supports continuation of the trial because $\epsilon_2/(1-\epsilon_1) < T_1(d_i) < (1-\epsilon_2)/\epsilon_1$. Alternatively if the test supports H_0, ie $T_1(d_i) < \epsilon_2/(1-\epsilon_1)$ then the decision might be that no more resources need to be spent with this agent at the current dose level.

As a next step, we simultaneously test for efficacy ie, $H_0 : Q(d_i) \leq q_0$ against $H_1 : Q(d_i) \geq q_1$ and at the same time we want to test for adequate toxicity rate, ie, $H_0 : R(d_i) > s_0$ against $H_1 : R(d_i) \leq s_1$, $0 < s_1 \leq s_0 < 1$; where $R(d_i)$ denotes the true toxicity rate at dose d_i. Let A_1, A_0 denote the restricted space for a under H_1, H_0, and B_1, B_0 denote the space for b under H_1, H_0. Given $\Omega(j) = \{(x_1, y_1, v_1), ..., (x_j, y_j, v_j)\}$, and assuming j^* patients have been treated at level d_i, let

$$\mathcal{H}_{\{A,B\}} = \int_B \int_A \prod_{l=1}^{j^*} \beta_i^{bv_l}(1-\beta_i^b)^{(1-v_l)} \prod_{l=1}^{j^*} \alpha_i^{ay_l}(1-\alpha_i^a)^{(1-y_l)} g_1(a) g_2(b)\, da\, db \tag{5}$$

where $g_1(a), g_2(b)$ denote pre-specified probability distribution functions for the parameters a and b respectively. The sequential test is defined as the ratio of the functions $\mathcal{H}$ when integrated within the respective regions under H_1, H_0 respectively and it is given by $T(d_i) = \mathcal{H}_{\{A_1,B_1\}}/\sum_{\mathcal{R}} \mathcal{H}_{\{A_{\mathcal{R}},B_{\mathcal{R}}\}}$ where $\mathcal{R}$ denotes different regions for the parameters a and b. The regions under the null correspond to different clinical hypothesis and one can modify the region under the null as well as the indifference regions accordingly. As a special case, if we want to test efficacious and non toxic dose against non-efficacious and toxic dose, then $H_0 : b \geq b_0$ and $a \leq a_0$ against $H_1 : b \leq b_1$ and $a \geq a_1$ and the test equals to $[\mathcal{H}_{\{A_0,B_0\}}]^{-1}\mathcal{H}_{\{A_1,B_1\}}$. Specifically, the test at level d_i is given by:

$$T_2(d_i) = \frac{\int_{B_1}\int_{A_1}\prod_{l=1}^{j^*}\beta_i^{bv_l}(1-\beta_i^b)^{(1-v_l)}\prod_{l=1}^{j^*}\alpha_i^{ay_l}(1-\alpha_i^a)^{(1-y_l)}g_1(a)g_2(b)\,da\,db}{\int_{B_0}\int_{A_0}\prod_{l=1}^{j^*}\beta_i^{bv_l}(1-\beta_i^b)^{(1-v_l)}\prod_{l=1}^{j^*}\alpha_i^{ay_l}(1-\alpha_i^a)^{(1-y_l)}g_1(a)g_2(b)\,da\,db}$$

where $A_0 : (0, a_0)$, $A_1 : (a_1, \infty)$ and $B_0 : (b_0, \infty)$ and $B_1 : (0, b_1)$ are the corresponding regions for the hypothesis given above. Alternatively, we can choose to test for adequate efficacy rate at level i assuming the value of the toxicity rate is known. The hypothesis will then be $H_0 : b \geq b_0(a)$ against $H_1 : b \leq b_1(a)$ conditional on a known value for a, for example the current estimate of $\hat{a}$. For simplicity we can use the mean of a or the maximum likelihood estimate of $\hat{a}$ (maximum) as a plug in estimate in Equation 5. In the context of Phase I designs, given the small sample size involved, the operating characteristics provide our main guide to the practical usefulness of such approximations.

In practice, we might approximate the composite hypotheses by using simple point hypotheses rather than composite ones. Integrating over a composite hypothe-

sis amounts to taking a mean, therefore we can approximate this mean (of a function) by the same function of the mean. Assume that at the current level d_i where patient j is being treated we want to test the hypotheses: $H_0 : Q(d_i) = q_0$ against $H_1 : Q(d_i) = q_1$, where q_0, q_1 denote low and desirable efficacy rates respectively. We can calculate the sequential test after j patients have been accrued at level d_i given by:

$$T_3(d_i) = r_i(j) \log \left(\frac{q_1(1-q_0)}{q_0(1-q_1)} \right) + j \log \frac{(1-q_1)}{(1-q_0)} \tag{6}$$

where $r_i(j)$ is the sum of responders treated at d_i who also have efficacy response measured. This test uses the empirical estimate of the number of efficacy responses assuming binomial distribution (O'Quigley et al. 2001). These tests help us decide whether the current dose is efficacious (T_1 or T_3), or efficacious and safe (T_2) which are the secondary criteria used to define the RP2D.

2.3 Theoretical Properties

The Average Sample Number (ASN) and Operating Characteristic (OC) function of the proposed tests are of interest both from a theoretical and applied perspective. Let us denote the probability that the sequential process will terminate with the acceptance of H_0, at dose level i when b is the true value of the parameter as $L_i(b)$. Let $E_1 = (1-\epsilon_2)/\epsilon_1$ and $E_2 = \epsilon_2/(1-\epsilon_1)$ be the boundaries for the sequential test T_1. The following lemma is needed for the proof of Theorem 1.

Lemma 1 *The Operating Characteristic function $L_i(b)$ is given by*
$L_i(b) = [E_1^{h(b)} - 1]/[E_1^{h(b)} - E_2^{h(b)}]$ *where for any chosen $b > 0$, $h(b)$ is the solution of*

$b = \left[\log\left(c^*\left(b_1, b_0, h\right) - 1\right) - \log\left(c^*(b_1, b_0, h) - \beta_i^{(b_1 - b_0)h}\right)\right]\left[1/\left(\log(\beta_i)\right)\right]$, *where* $c^*(b_1, b_0, h) = [(1 - \beta_i^{b_1})/(1 - \beta_i^{b_0})]^h$.

Proof: See Appendix A.1.

Theorem 1 *The expected value of* T_1 *is given by*

$$\mathbb{E}(T_1(d_i)) = L_i(b)E_2 + (1 - L_i(b))E_1.$$

Proof: See Appendix A.1 in Supplementary Materials.

If we assume that all patients in the expansion cohort are treated at the same level, say d_m then the probability to terminate the clinical trial overall will be equal to the probability of terminating the experiment at d_m. However, because experimentation occurs at two levels, the probability of terminating the clinical trial in favor of H_0, $L^*(b)$ is given by $L^*(b) = \sum_{i=1}^{k} \Pr(\text{stop for } H_0|d_i)\pi(d_i) = \sum_{i=1}^{k} L_i(b)\pi(d_i)$ where, $\pi(d_i)$ corresponds to the true probability of experimenting at level d_i. It has been shown [O'Quigley 2006] that under certain conditions, CRM has the property to converge to the true MTD, say d_{m^*} and that $\pi(d_i) \to 0, i \neq m^*$, while $\pi(d_i) \to 1, \text{for } i = m^*$, as N increases. In the proposed design we assume that experimentation will focus at two levels around d_{m^*} as defined in Section 2.1 and the corresponding probabilities will depend on the randomization probabilities accordingly. Thus there exist two levels whose $\pi(d_i) \neq 0$, specifically $\pi(d_{m^*}) > 0$ and $\pi(d_{m^*+1}) > 0$, while for the remaining levels we assume $\pi(d_i) \to 0, i \neq m^*, m^* + 1$. If J, and therefore N, increase without bound, we could approximate these probabilities with the stable distribution of patients included at d_{m^*}. Thus, $L^*(b) = \pi(d_{m^*})L_{m^*}(b) + [1 - \pi(d_{m^*})]L_{m^*+1}(b)$.

Numerical approximations for $L_i(b)$ (Supplementary Materials, Appendix A.2) and the expected value of n, which is the number of patients required to reach a decision in favor of H_1 are derived in Appendix B.

3 Applications

3.1 An example of a clinical trial in solid tumors

For illustration we use a published Phase I trial that followed the 3+3 algorithm during the dose escalation phase followed by an expansion cohort [Isambert et al. 2012]. The trial was in patients with advanced solid tumors with metastatic or nonresectable cancer. The aim of the study was to find the RP2D of aflibercept when in combination with docetaxel. The trial followed the 3+3 design with modifications and recommended level 4 as the MTD, based on two adverse events of hypertension observed in levels 5 and 6 that met the definition of DLT. The dose escalation phase consisted of 34 patients, ten of whom were treated at the MTD. An additional 20 patients were treated at the MTD, as part of a DEC for a total of 54 (34+20) patients in the trial. The observed DLT rates at each one of the six dose levels that were included in the trial before the DEC were 1/7, 0/3, 0/6, 0/10, 1/5, 1/3 respectively. We illustrate how we could allocate patients to levels during the expansion phase using a model based algorithm sequentially after each patient's response is updated. Assume hypothetically that during the expansion phase, the 20 additional patients were randomized as described in Section 2.1. The toxicity outcomes of the additional patients were simulated using isotonic regression estimates of the observed rates [**?**] obtained during the dose escalation phase. Using this algorithm, 11 patients would

have been treated at level 5 and 9 patients at level 6 since the estimated DLT rate indicated that the MTD lies between levels 5-6. At the end of the study the predicted probabilities of toxicity at each level are equal to 0.01, 0.03, 0.07, 0.13, 0.22, 0.32 for a skeleton of $\alpha_i = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$, and $\hat{a} = 2.2175$, and hence the RP2D is dose 5 for a target toxicity rate of 25% (Table 1). As a next step, we assume that we have efficacy measured on a subset of patients as denoted by J, and we update $\hat{a}$ and $\hat{b}$ (Equation 2) sequentially. The algorithm without randomization allocated 9 patients to level 5 and 11 patients to level 6, and the observed efficacy rate at level 5 and 6 was 0/9 and 5/11 respectively. The test given by Equation 6 recommends terminating the trial after 10 patients and supports adequate efficacy rate of 30% or more at level 6. Note that the recommendation at the beginning of the DEC is to continue the trial, then it switches to evidence favoring H_0 as more patients are treated at level 5 without efficacy responses and then by patient number 10, it reaches a decision in favor of a response rate of 30% for dose level 6. Note that by treating all 20 patients at the current MTD, the established MTD was dose level 4.

3.2 Designing and running a trial

Logistically calculating the test statistic after each patient inclusion is challenging. Instead, we can provide investigators in advance the number of efficacy responses required to make a decision in favor of accepting or rejecting the null hypothesis respectively and include these numbers in the protocol as shown in Table 2. Investigators need to only count the number of efficacy responses $r_i(j)$; if this number lies in the the interval $(r_1^*(j), r_2^*(j))$ we continue experimentation, if $r_i(j) \geq r_2^*(j)$ we conclude in favor of H_1 and if $r_i(j) \leq r_1^*(j)$ we conclude in favor of H_0. An R code can

be provided by the first author to obtain the numbers $r_1^*(j), r_2^*(j)$ for pre specified values for Type I, II errors, and rates q_0, q_1. Table 2 shows that the model based test reaches a conclusion in favor of the alternative hypothesis faster. Selecting the values of these errors in an actual trial will depend on the operating characteristics of the SPRT which depend on the unknown and true, underlying parameters such as the difference in the efficacy rates. The results given in Table 4 (ASN) will help us decide the values of Type I and II errors. Type I and II errors set as high as 20% is not uncommon in DEC.

4 Simulation study

4.1 Operating Characteristics

In the simulation study we assume investigators follow the 3+3 design during the dose escalation phase, as this is typical in Phase I trials in oncology, followed by a DEC of additional J patients. The dose expansion phase is guided by the model following the completion of the 3+3. We assume that the trial established an initial estimate of the MTD using the 3+3 design [Iasonos et al. 2008]. Additional data, such as efficacy or PK response are obtained during the expansion phase. The DEC could be accrued under three different schemes as follows: Scheme 1: all patients are treated at the MTD established during the dose escalation phase (3+3). Scheme 2: after the MTD has been established using the 3+3 algorithm, patients accrued during the expansion phase are being randomized to two levels sequentially as m in Equation 3, might change at each step, based on the toxicity data from all patients as described in Section 2.1. Scheme 3: same as scheme 2, i.e., the DEC is being allocated to one

or two levels based on randomization, but the predicted probabilities of DLT are obtained using a bivariate outcome of toxicity and efficacy simultaneously utilizing all available information from all patients. In this last set of trials we calculated the sequential tests after each patient, using different forms of hypothesis testing as shown in Section 2.2.

The parameters used in the simulation study are as follows:

1. The true toxicity and efficacy rates at each dose level are denoted as R_i and Q_i respectively and are shown in Table 3. Scenario 1 denotes a case where there exists a safe dose which is not efficacious and the efficacious level is not safe; in Scenario 2, d_5 is the dose which is safe and simultaneously efficacious. Scenario 3 has two levels that are safe but only one is efficacious. Scenario 4 is used as a theoretical bound of how well the method is doing since here the models are following the true rates.

2. The skeleton values for toxicity and efficacy for the parameters α_i, β_i are given by skeleton values, α_i equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 for the 6 levels respectively.

3. The sample size during the 3+3 stage varies since the trial stops at any time after observing 2/6 DLTs and after 6 patients have been treated at the level below and with at most 1/6 DLTs at the MTD. The sample size for the DEC J, varied from 12, 25, 50 when evaluating accuracy of dose recommendation and percent of patients treated. A maximum sample size of 200 is used in order to estimate the average sample number.

4. Type I and Type II errors, ϵ_1, ϵ_2 are set at 10%.

5. For the hypothesis test involved in Section 2.2 the efficacy rates that are considered too low and desirable are $q_0 = 10\%$ and $q_1 = 30\%$ respectively. The acceptable threshold for toxicity was set at $s_1 = s_0 = \theta = 30\%$. Alternatively one can test the hypotheses $H_0 : R(d_i) > s_0 = 0.3$ and $H_1 : R(d_i) \leq s_1 = 0.2$ so that we can reject levels with DLT rate > 0.3 while considering levels with rates ranging from 0.2-0.3 as supporting further experimentation ($s_1 \leq \theta \leq s_0$). The choice of values of s_1, s_0 depend on the individual clinical scenario.

6. The uniform distribution was used for the distribution of $g_1(a)$ and $g_2(b)$ in the test statistics T_1, T_2 with a sensitivity analysis using Gamma distribution with different parameters allowing for larger variance.

Table 3 provides a summary of the dose recommendation across many simulated trials. We see that following a model based approach during the expansion phase increases the accuracy of finding the true MTD. The increase is on average 40% in absolute percentage points or 35% improvement compared to assigning all patients at the MTD found during the dose escalation phase. The increase in accuracy is apparent across many scenarios regardless of the location of the MTD. This indicates that we ought to take into account the toxicity responses from the additional patients accrued during the expansion phase, and while the efficacy responses are considered secondary in terms of dose allocation, there is no loss in accuracy by estimating efficacy rates simultaneously to the toxicity rates. Updating these rates as the trial is ongoing and more data are accumulated seems the right approach, both from an ethical standpoint since we maintain safety as the primary objective, as well as in terms of efficiency in the dose allocation algorithm.

Table 4 shows the results from carrying out sequential probability ratio tests. Each test is calculated at the dose level where the current patient is being treated thus for each one of the three tests we have summarized the percent of trials where the decision was in favor of $H_1 : R(d_i) \geq 0.30$ at the MTD or at MTD plus or minus a level. The first scenario is a case where there exists a safe dose which is not efficacious, and the dose above this level is unsafe and efficacious. Table 4 shows that all tests correctly accept MTD+1 as an efficacious level, but fail to support MTD as efficacious level, given the fact that 20-24% of the trials decided in favor of H_1 when testing efficacy at MTD. Note that T_2 supports less frequently efficacy at MTD+1 since T_2 tests simultaneously for a safe and efficacious level, and MTD+1 fails these requirements. In Scenario 2, the tests correctly identify the MTD as efficacious and safe dose which is supported by the true rates while again T_2 often rejects MTD+1 based on safety compared to the other two tests. In scenario 3, the MTD falls between dose level 4 or 5 based on safety alone, whereas dose level 5 is more efficacious. Thus the tests are deciding in favor of H_1 at the MTD or/and MTD+1. The last scenario represents a case where the working models follow the true rates and the MTD is efficacious. In this scenario, 95% of trials reached a decision in favor of H_1.

The sample size required to make a decision in favor of H_1 or H_0 was calculated for each trial (denoted as ASN^*). In addition, we calculated the average sample number (ASN) required to make a decision in favor of H_1 calculated at each level separately. Figure 2 shows the distribution of ASN when the test is calculated at any level, regardless if that was the MTD. We see that depending on the scenario and whether there exists a safe and efficacious dose, then the tests either terminate in favor of H_1 early as in Scenario 2 or they take longer to reach a decision for H_1 as

in Scenario 1, when a level with high efficacy might not be a safe dose. The median sample size is shown in Table 4 and it is substantially smaller with T_2 as compared to the test based on the empirical rates. This supports increase in efficiency by using complete data on toxicity and efficacy. Using a model based test, we need on average 20 patients to reach a decision in favor of H_1 which is consistent with current, clinical practice. Additional simulations with a scenario where the efficacious dose was highly toxic (R =50%) showed that each test takes longer (larger ASN) to reach a decision in favor of H_1 as there is little experimentation at the efficacious level due to safety concerns and the safe levels are not efficacious.

4.2 Sensitivity analysis

In order to assess whether Type I and Type II errors are controlled and what is the effect of small sample sizes in the operating characteristics, we report the decision reached by the three sequential tests when a sample is fixed at 12, 25, 50 (Table 5) in the context of dose expansion cohorts. The number of trials in favor of H_1 in Scenario 1 is an estimate of the Type I error since the MTD is not efficacious (response rate at MTD is 10%),and we see Type I error is less than the set value of 10%. In Scenario 2 the MTD has an efficacy rate of 30%, so the number of trials in favor of H_0 is an estimate of the Type II error and again it is close to the set value of 10%. A sensitivity analysis for the choice of distribution $g(b)$ can inform us about the influence of the choice of such distribution. In our sensitivity analysis for the distributions used in T_1,T_2 we assumed Gamma distribution with various shape and scale parameters that result in various mean and variance values. The results show that in certain cases, Gamma with small variance can be more informative

than Uniform since the sequential test was reaching a decision early on and a higher number of trials were reaching a decision in favor of H_1 (Table 6). This indicated that the prior was informative when the variance was small (1.25) but was not informative with larger variance (variance of 4 for a and 6 for b). The relative importance of the prior depends on the sample size. Since here we are not only dealing with estimation and finding the location of the MTD but in addition we are using sequential tests to guide a decision process of whether to terminate or continue the trial, then the influence of the prior matters early on in the trial. If the decision is to stop the trial early, then we will never increase N to recover from that decision. For these reasons we suggest using a non informative distribution such as Uniform.

We assessed the robustness of our proposed approach to the assumption of conditional independence by running additional simulations under various values of correlation parameter(Appendix C of Supplementary Materials).

5 Conclusion

In this paper, we outline methodology to adequately design and monitor DEC in Phase I trials in oncology. This methodology provides initial estimates of efficacy activity, as this is one of the goals of DEC, and aids in providing a go/no go decision rule at the current dose. It is meant to serve as a guide to help clinical investigators decide whether the dose with the best chance for efficacy activity should be selected or whether more patients need to be treated and at which level; versus abandonment of the investigational drug altogether. Given that there is still considerable uncertainty in the selection of the MTD after the dose escalation part of the phase I trial, we

propose experimentation at more than a single level during the expansion phase in order to estimate efficacy at more than one dose level. This dose exploration during the expansion phase will provide support to the decision of whether a higher or lower dose is needed with regards to efficacy. The secondary efficacy measure can capture evidence of biomarker expression or tumor absorption, which allows the dose expansion to focus experimentation in this targeted population and help select the appropriate patient population for future trials. Finally, an important feature of the proposed approach is that it does not require toxicity and efficacy endpoints to be obtained at the same time. The sequential equations can be updated at any point, as data become available, making this approach logistically simple to implement.

Sequential tests are efficient in terms of sample size requirements in order to test for a specific hypothesis [Wald 1947, Cheung 2007] and they have been used extensively in other contexts such as randomized Phase II or III studies with multiple looks. In the scenarios presented above, 35 patients would be required to test for a response rate of 10% versus 30% with Type I and II errors set at 10% using Simon's optimal design. In the simulation study we presented, when there exists a safe and efficacious dose, on average 16-24 patients are required to terminate the study. Depending on the location of the MTD and the steepness of the efficacy curve, the sample size savings can be as high as 20 patients. This is a considerable amount in patient savings given that at the end of the study we simultaneously address two objectives: establish the dose and estimate the efficacy associated with the proposed dose. In this paper, we followed clinical practice by adding the DEC following the 3+3 design. It would be preferable to use a model based design from the beginning of the trial and to include the DEC in the model estimation. In such scenario no

additional methodological considerations are needed. However, the choice of which design to use is up to clinical investigators. Regardless of how we obtain the data, SPRT allows us to select a dose based on promising activity, since it allows us to stop at any point for futility; stop early if there is a strong efficacy signal or suggest that further experimentation is needed if the results are inconclusive. Sequential tests make an efficient use of the data in the context of Phase I trials and this is the best we can do under such small size studies before committing ourselves to embark on a larger Phase II or a Phase III study.

Supplementary Materials: The reader is referred to the on-line Supplementary Materials for the proofs of Lemma 1, Theorem 1 (Appendix A.1), approximations of the composite hypotheses (Appendix A.2) and Average Sample Number (Appendix B); and further simulations as part of a sensitivity analysis (Appendix C).

Funding:This work was partially supported by National Institute of Health (Grant Number 1R01CA142859).

References

Cheung YK. Sequential implementation of stepwise procedures for identifying the maximum tolerated dose. Journal of the American Statistical Association. 2007; 102, 1448-1461.

Iasonos A, O'Quigley J. Design considerations for dose-expansion cohorts in phase I trials. *Journal of Clinical Oncology* 2013; Nov 1;31(31):4014-21.

Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. A comprehensive comparison of the continual reassessment method to the standard 3 + 3 dose escalation scheme in Phase I dose-finding studies. *Clinical Trials* 2008; 5:465-77.

Isambert N FG, Zanetta S, You B, et al. Phase I dose-escalation study of intravenous aflibercept in combination with docetaxel in patients with advanced solid tumors. Clin Cancer Res 2012;18(6):1743-50.

Lee SM, Cheung YK. Calibration of prior variance in the Bayesian continual reassessment method. Stat Med. 2011 Jul 30;30(17):2081-9.

Manji A, Brana I, Amir E, et al. Evolution of clinical trial design in early drug development: systematic review of expansion cohort use in single-agent phase I cancer trials. J Clin Oncol. 2013 Nov 20;31(33):4260-7.

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

O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach.*Biometrics* 1996 ;52:673-84.

O'Quigley J, Conaway M. Continual Reassessment and related dose finding designs. Statistical Science. 2010. Vol 25. No 2, 202-216.

O'Quigley J, Hughes MD, Fenton T.Dose-finding designs for HIV studies. Biometrics. 2001 Dec;57(4):1018-29.

O'Quigley J. Theoretical study of the continual reassessment method. *Journal of Statistical Planning and Inference* 136 (2006) 1765 - 1780.

O'Quigley J, Hughes MD, Fenton T, Pei L. Dynamic calibration of pharmacokinetic parameters in dose-finding studies. Biostatistics (2010) 11 (3): 537-545.

Pei L and Hughes MD. A Statistical Framework for Quantile Equivalence Clinical Trials with Application to Pharmacokinetic Studies that Bridge from HIV-Infected Adults to Children. Biometrics.2008. Volume 64, Issue 4, pages 11171125.

Rogatko A, Schoeneck D, Jonas W, et al. Translation of innovative designs into phase I trials. J Clin Oncol 2007;25(31):4982-6.

Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012;366(26):2443-54.

Wald A. Sequential Analysis. John Wiley and Sons. New York. 1947.

Yuan Y. Bayesian Phase I/II adaptively randomized oncology trials with combined drugs. The Annals of Applied Statistics. 2011; 5, 924-942.

Table 1: Illustrative Trial: Parameter estimates and decision rules for a hypothetical expansion cohort of 20 patients enrolled in addition to the 34 patients accrued in the dose escalation trial. Type I and II errors ar set at 20% and efficacy rates are set at $q_0 = 5\%$, and $q_1 = 30\%$ under H_0, H_1 respectively.

Patient No.	Data: (d_i, y)	$\hat{a}$	Data: (d_i, y, v)	$\hat{a}$	$T_3(d_i)$	Decision
35	6 0	2.1857	5 0 0	2.1857	-0.31	continue
36	6 1	2.2512	5 0 0	2.2369	-0.61	continue
37	6 0	2.1221	5 0 0	2.2868	-0.92	continue
38	6 0	2.1824	5 1 0	2.3355	-1.22	continue
39	6 0	2.2419	5 0 0	2.1611	-1.53	acc H_0
40	6 0	2.3004	5 0 0	2.2057	-1.83	acc H_0
41	5 0	2.3580	5 0 0	2.2493	-2.14	acc H_0
42	6 0	2.4013	5 0 0	2.2919	-2.44	acc H_0
43	5 1	2.4569	5 0 0	2.3336	-2.75	acc H_0
44	5 1	2.2869	6 0 1	2.3743	1.79	rej H_0 /stop
45	5 0	2.1453	6 0 1	2.4264	3.58	rej H_0
46	6 1	2.1829	6 0 1	2.4778	5.38	rej H_0
47	5 1	2.0913	6 1 0	2.5292	5.07	rej H_0
48	5 0	1.9817	6 0 0	2.4050	4.76	rej H_0
49	5 0	2.0160	6 0 0	2.4519	4.46	rej H_0
50	5 0	2.0497	6 0 0	2.4987	4.15	rej H_0
51	5 0	2.0827	6 0 0	2.5456	3.85	rej H_0
52	5 0	2.1152	6 1 1	2.5902	5.64	rej H_0
53	5 0	2.1470	6 0 1	2.4741	7.43	rej H_0
54	6 0	2.1783	6 0 0	2.5169	7.13	rej H_0
55		2.2175		2.5588		

Table 2: Acceptance and rejection numbers used for trial design, $(r_1^*(j), r_2^*(j))$ for the test statistics given in Equations 4 and 6.

	$\epsilon_1 = 0.20, \epsilon_2 = 0.20$ $q_0 = 0.05, q_1 = 0.30$ $r_1^*(j), r_2^*(j)$	$\epsilon_1 = 0.20, \epsilon_2 = 0.20$ $q_0 = 0.15, q_1 = 0.30$ $r_1^*(j), r_2^*(j)$
Patient No.	for T_1 for T_3	for T_1 for T_3
1	(- , -) (- , 1)	(- , -) (- , 3)
2	(- , -) (- , 1)	(- , 2) (- , 3)
3	(- , 2) (- , 2)	(- , 2) (- , 4)
4	(- , 2) (- , 2)	(- , 3) (- , 4)
5	(- , 2) (0 , 2)	(- , 3) (- , 4)
6	(- , 2) (0 , 2)	(- , 3) (- , 4)
7	(- , 2) (0 , 2)	(- , 4) (- , 5)
8	(1 , 2) (0 , 2)	(1 , 4) (- , 5)
9	(1 , 3) (0 , 2)	(1 , 4) (- , 5)
10	(1 , 3) (0 , 3)	(1 , 4) (- , 5)
11	(1 , 3) (0 , 3)	(1 , 5) (- , 5)
12	(1 , 3) (1 , 3)	(1 , 5) (0 , 6)
13	(1 , 3) (1 , 3)	(2 , 5) (0 , 6)
14	(1 , 3) (1 , 3)	(2 , 5) (0 , 6)
15	(2 , 4) (1 , 3)	(2 , 6) (0 , 6)
16	(2 , 4) (1 , 3)	(2 , 6) (1 , 6)
17	(2 , 4) (1 , 4)	(2 , 6) (1 , 7)
18	(2 , 4) (1 , 4)	(2 , 6) (1 , 7)
19	(2 , 4) (2 , 4)	(3 , 7) (1 , 7)
20	(2 , 4) (2 , 4)	(3 , 7) (1 , 7)

Table 3: Proportion of trials recommending each dose level for the four scenarios for $J = 50$. Scheme 1: 3+3 design followed by expansion; Scheme 2: expansion is guided by CRM and randomization; Scheme 3: expansion is guided by CRM and randomization using efficacy as well as toxicity. True toxicity and efficacy rates used in the simulation study are denoted with R, Q respectively.

Scenario	Levels:	d_0	d_1	d_2	d_3	d_4	d_5	d_6
1	R		0.10	0.15	0.30	0.45	0.50	0.60
	Q		0.05	0.09	0.10	0.30	0.40	0.45
	Scheme 1	0.08	0.19	0.40	**0.26**	0.06	0.01	
	% patients		0.25	0.38	**0.25**	0.07	0.01	0.00
	Scheme 2			0.11	**0.61**	0.20	0.01	
	% patients		0.10	0.18	**0.45**	0.20	0.03	0.00
	Scheme 3			0.12	**0.61**	0.18	0.01	
	% patients		0.10	0.18	**0.44**	0.20	0.03	0.00
2	R		0.05	0.10	0.15	0.20	0.30	0.60
	Q		0.01	0.05	0.09	0.10	0.30	0.40
	Scheme 1	0.03	0.09	0.17	0.23	0.26	**0.20**	0.01
	% patients		0.14	0.19	0.23	0.23	**0.17**	0.02
	Scheme 2		0.00	0.00	0.02	0.26	**0.63**	0.07
	% patients		0.07	0.07	0.11	0.25	**0.41**	0.08
	Scheme 3		0.00	0.00	0.03	0.28	**0.59**	0.07
	% patients		0.07	0.07	0.11	0.26	**0.40**	0.08
3	R		0.05	0.1	0.15	0.27	0.33	0.60
	Q		0.01	0.05	0.09	0.10	0.30	0.40
	Scheme 1		0.10	0.17	0.36	**0.20**	**0.13**	0.02
	% patients		0.14	0.19	0.33	**0.19**	**0.11**	0.02
	Scheme 2		0.00	0.00	0.06	**0.45**	**0.43**	0.04
	% patients		0.06	0.07	0.14	**0.34**	**0.31**	0.06
	Scheme 3		0.00	0.0	0.08	**0.43**	**0.43**	0.04
	% patients		0.06	0.07	0.15	**0.34**	**0.31**	0.05
4	R		0.02	0.06	0.12	0.20	0.30	0.41
	Q		0.02	0.06	0.12	0.20	0.30	0.41
	Scheme 1		0.05	0.13	0.28	0.27	**0.20**	0.08
	% patients		0.09	0.15	0.27	0.24	**0.17**	0.07
	Scheme 2		0.00	0.00	0.01	0.20	**0.61**	0.18
	% patients		0.02	0.06	0.09	0.22	**0.42**	0.17
	Scheme 3		0.00	0.00	0.01	0.21	**0.60**	0.18
	% patients		0.05	0.06	0.09	0.22	**0.41**	0.17

Table 4: Proportion of trials deciding in favor of H_1 when including patients treated at the MTD, MTD -1, MTD +1 for the four scenarios of Table 3 and trials under Scheme 3. T_1 is based on model based efficacy rates; T_2 is based on an acceptable region of efficacy and toxicity and T_3 is based on the observed efficacy rate. Median Sample size and interquartile range (IQR) in order to make a decision in favor of H_1 at any dose.

	Test :	MTD-1	MTD	MTD+1	Median n (IQR)
Scenario 1					
	T_1	0.06	0.20	0.56	35 (9, 201)
	T_2	0.18	0.26	0.49	34 (11, 201)
	T_3	0.03	0.17	0.52	60 (17, 201)
Scenario 2					
	T_1	0.11	0.84	0.61	17 (7, 40)
	T_2	0.25	0.85	0.49	16 (8, 36)
	T_3	0.07	0.84	0.56	23 (12, 50)
Scenario 3					
	T_1	0.07	0.53	0.60	24 (8, 71)
	T_2	0.19	0.58	0.54	21 (9, 58)
	T_3	0.04	0.50	0.57	33 (13, 89)
Scenario 4					
	T_1	0.43	0.95	0.64	10 (5,21)
	T_2	0.53	0.96	0.62	11 (6,19)
	T_3	0.39	0.94	0.61	15 (9,27)

Table 5: Proportion of trials deciding in favor of H_1, H_0, or inconclusive when including patients treated at the MTD under different sample size requirements accrued during the expansion cohort (J = 12,25,50). The four scenarios represent trials simulated under Scheme 3 of Table 3.

		Scenario 1			Scenario 2		
Scenario 1	J	H_1	H_0	continue	H_1	H_0	continue
T_1	50	0.06	0.80	0.13	0.71	0.11	0.17
	25	0.06	0.68	0.26	0.55	0.10	0.34
	12	0.05	0.48	0.46	0.37	0.11	0.52
T_2	50	0.07	0.77	0.16	0.73	0.11	0.17
	25	0.08	0.62	0.30	0.60	0.09	0.31
	12	0.09	0.39	0.52	0.45	0.12	0.43
T_3	50	0.05	0.72	0.23	0.68	0.09	0.24
	25	0.04	0.49	0.47	0.49	0.05	0.45
	12	0.02	0.21	0.77	0.29	0.01	0.69
			Scenario 3			Scenario 4	
	J						
T_1	50	0.39	0.45	0.15	0.77	0.05	0.18
	25	0.33	0.37	0.30	0.58	0.06	0.36
	12	0.25	0.23	0.52	0.37	0.09	0.54
T_2	50	0.39	0.44	0.16	0.79	0.04	0.17
	25	0.34	0.34	0.33	0.63	0.08	0.29
	12	0.31	0.20	0.49	0.47	0.12	0.41
T_3	50	0.37	0.40	0.23	0.75	0.02	0.23
	25	0.28	0.27	0.45	0.54	0.02	0.44
	12	0.15	0.09	0.76	0.30	0.01	0.69

Table 6: Proportion of trials deciding in favor of H_1 when including patients treated at the MTD, MTD -1, MTD +1 for the four scenarios when using different prior distributions. T_1 is based on model based efficacy rates; T_2 is based on an acceptable region of efficacy and toxicity. Median Sample size (ASN) to make a decision in favor of H_1 at any dose. G : Gamma with parameters (scale,shape).

$g(b)$ in T_1					$g_2(b), g_1(a)$ in T_2				
Scenario 1	MTD-1	MTD	MTD+1	ASN	Scenario 1	MTD-1	MTD	MTD+1	ASN
Uniform	0.06	0.20	0.56	35	Uniform	0.18	0.26	0.49	34
G(1.5,2)	0.11	0.26	0.59	14	G(1.5,2);G(1,2)	0.17	0.26	0.49	34
G(2,1)	0.15	0.31	0.63	12	G(2,1);G(2,1)	0.17	0.25	0.48	36
G(5, 0.5)	0.49	0.74	0.78	4	G(5, 0.5);G(4,0.5)	0.20	0.29	0.50	30
Scenario 2					Scenario 2				
Uniform	0.11	0.84	0.61	17	Uniform	0.25	0.85	0.49	16
G(1.5,2)	0.15	0.84	0.61	12	G(1.5,2);G(1,2)	0.23	0.85	0.48	16
G(2,1)	0.19	0.85	0.61	11	G(2,1);G(2,1)	0.22	0.84	0.48	16
G(5, 0.5)	0.50	0.94	0.72	4	G(5, 0.5);G(4,0.5)	0.26	0.86	0.49	16
Scenario 3					Scenario 3				
Uniform	0.07	0.53	0.60	24	Uniform	0.19	0.58	0.54	21
G(1.5,2)	0.10	0.54	0.61	15	G(1.5,2);G(1,2)	0.17	0.58	0.54	21
G(2,1)	0.14	0.58	0.62	13	G(2,1);G(2,1)	0.17	0.57	0.53	22
G(5, 0.5)	0.45	0.79	0.72	4	G(5, 0.5);G(4,0.5)	0.20	0.59	0.54	20
Scenario 4					Scenario 4				
Uniform	0.43	0.95	0.64	10	Uniform	0.53	0.96	0.62	11
G(1.5,2)	0.44	0.94	0.66	6	G(1.5,2);G(1,2)	0.52	0.96	0.62	11
G(2,1)	0.49	0.95	0.64	6	G(2,1);G(2,1)	0.52	0.96	0.61	11
G(5, 0.5)	0.69	0.99	0.70	3	G(5, 0.5);G(4,1/2)	0.54	0.97	0.62	11

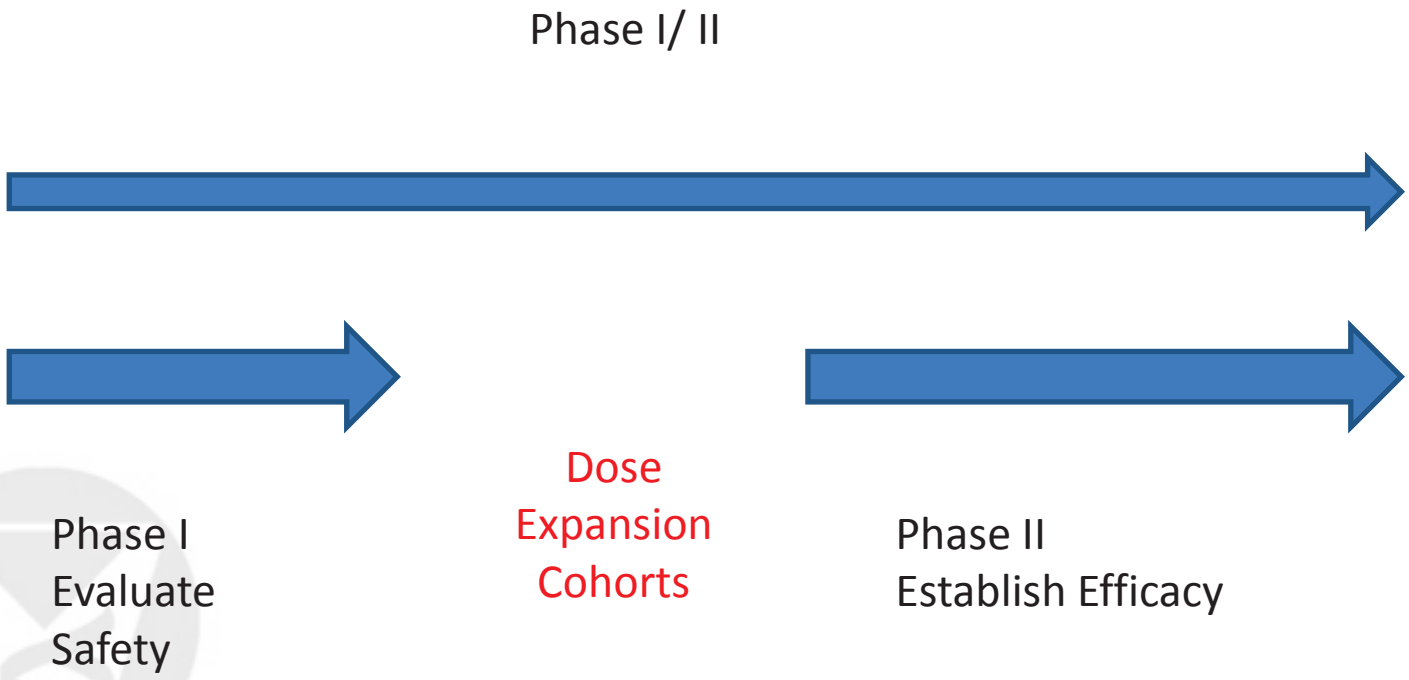

Figure 1: Illustrating the use of dose expansion cohorts following a dose escalation study.

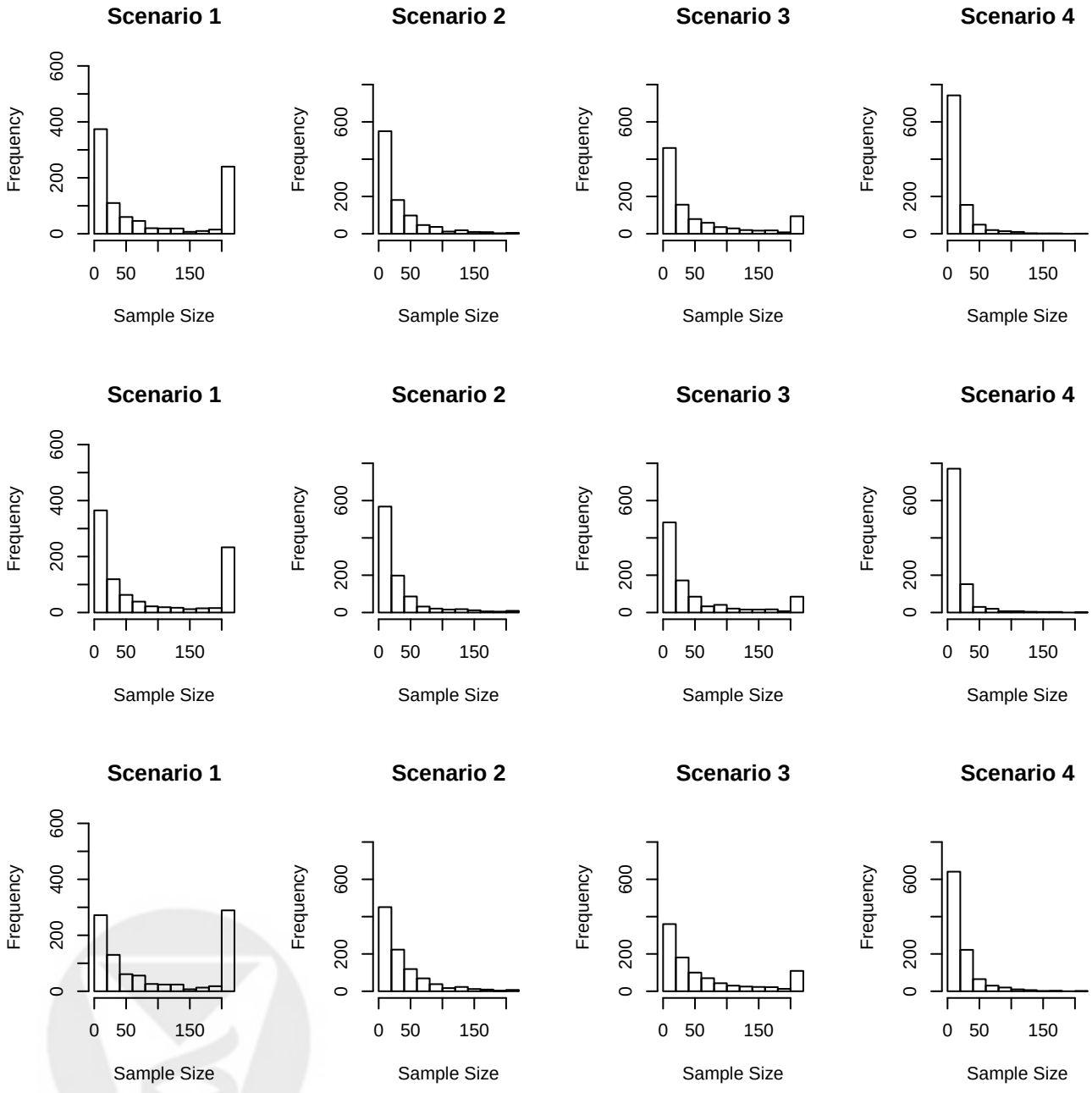

Figure 2: Distribution of sample size required to make a decision in favor of H_1 at any level regardless if that was the MTD. Horizontal panels show sequential tests based on T_1, T_2, T_3 for the 4 scenarios respectively.