HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2018 January 30.

Published in final edited form as:

Stat Med. 2017 January 30; 36(2): 204–214. doi:10.1002/sim.6894.

Sequential Monitoring of Phase I Dose expansion cohorts

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 relatively recent development in the design of Phase I dose finding studies is the inclusion of expansion cohort(s), that is the inclusion of several more patients at a level considered to be the maximum tolerated dose (MTD) established at the conclusion of the "pure" Phase I part. Little attention has been given to the additional statistical analysis, including design considerations, that we might wish to consider for this more involved design. For instance how can we best make use of new information that may confirm or may tend to contradict the estimate of the MTD based on the dose-escalation phase. Those patients included during the dose expansion phase may possess different eligibility criteria. During the expansion phase we will also wish to have an eye on any evidence of efficacy, an aspect that clearly distinguishes such studies from the classical Phase I study. Here we present methodology that enables us to continue the monitoring of safety in the DEC while simultaneously trying to assess efficacy and, in particular, which disease types may be the most promising to take forward for further study. The most elementary problem is where we only wish to take account of further toxicity information obtained during the DEC, and where the initial design was model based or the standard 3+3. More complex set-ups also involve efficacy and the presence of subgroups.

Keywords

dose finding; Phase I trials; dose expansion; sequential test; sequential probability ratio test

1 Background and motivation

Phase I protocols with dose expansions can be viewed as having two phases: a dose escalation phase that aims to estimate the maximum tolerated dose (MTD) and then, a dose expansion phase that has several objectives. The inclusion of a dose expansion cohort (DEC) in a Phase I dose finding study introduces features that have an impact on both the design and the analysis. In the model based setting, where a model was used to guide the dose escalation, it can be argued that there is no real problem to address; we simply continue to recruit patients allocating the additional patients of the DEC to the current estimate of the MTD. However, such a view completely misses the objective of DEC and the clinical implications that make DEC unique. Although we continue to recruit patients in the DEC allocating the additional patients of the DEC to the current estimate of the MTD, we are now focused in a much narrower therapeutic window as we do experiment with fewer levels than what we started with. In other words, the dose escalation phase eliminates levels and provides an initial estimate of the MTD, while the DEC refines that initial estimate of the

MTD. Patients accrued during the DEC will often come from a different population having more narrowly defined characteristics, for example disease specific or histology specific cohorts. Furthermore, DEC are not looking at the outcomes in the same way and, in most cases, investigators wish to monitor the DEC not only for evidence of tolerability to the dose level but also for evidence of clinical or biological response. It is important then to recognize that Phase I trials with dose expansion cohorts are neither Phase I trials nor Phase I/II trials. [1, 2] Indeed, Phase I trials with DEC raise several questions relating to both the trial design and statistical analysis. In this paper, we consider methods that enable us to take into account of the information provided by the DEC.

The information obtained by the DEC may lead us to re-evaluate the recommended Phase II dose (RP2D) based on all of the information provided by the patients whether recruited during the DEC or during the first phase. We would like for designs to enable the estimation of efficacy in different settings in order to best guide our choice of which doses to take forward and for which groups into the Phase II context. Recent work on the methodology of Phase I designs has included techniques for dealing simultaneously with efficacy and toxicity [3, 4, 5]. It is of course much easier to adapt a model-based design for the Phase I part alone to a model based design for both the Phase I part together with the DEC. This is no doubt the best approach if, in advance, we are able to convince our clinical colleagues to relinquish the standard 3+3 for the first part. However, the clinicians often prefer to make use of the 3+3 design for the early part of the study, and yet are not opposed to a more sophisticated analysis for the DEC component. Previous reports have shown that Phase I trials use the 3+3 design approximately 90-95% of the time ([6]). Manji and colleagues showed that 68% of trials are multicenter and 75% of the trials in their review were industry sponsored among Phase I single agent trials examined in a 5 year period between 2006-2011. A dose escalation guided by the 3+3, followed by a DEC, does not mean however that a DEC, following a 3+3 design, cannot be analyzed. Although certainly less effective than working with a model based design from the outset, valuable inferences following the DEC, and during the DEC, can be made. Iasonos and O'Quigley (2013) [1] illustrated how a retrospective analysis of DEC data can be carried out and how the estimate of the MTD can be refined given the updated information.

The first phase of a Phase I trial with DEC has the goal of establishing the MTD. Once we have a sufficiently reliable estimate of the MTD we then embark on the expansion phase and this will involve an additional six to, sometimes more than, 50 patients [7]. The first phase typically defines the estimate of the MTD but the RP2D will be decided on the basis of the observed rates of toxicities, both before and following the DEC. The toxicity rates that are observed from the DEC will be taken into account by the clinical investigators and are not viewed as those that were used to drive estimation of the MTD. As a result the RP2D can be lower or higher than the MTD [8, 9] given all the information on safety together with the additional information on efficacy, pharmacokinetic studies or other factors. Whereas, for initial estimation of the MTD, only the rates of dose limiting toxicities at the different dose levels were taken into account, many other measurements, mostly concerned with efficacy, can be used for the expansion cohort that will help identify the most promising candidate for the RP2D. In many studies specific inclusion criteria such as the amount of disease or

information from biopsies can help focus the effort to identify areas of efficacy in subgroups of patients with specific biomarker status.

In attempting to articulate the precise aims of a Phase I trials with DEC we can see two broad areas; the first is whether or not to select an MTD based on the safety data alone, and the second is to carry out more studies for evidence of efficacy in particular patient subgroups while nonetheless controlling the rate of toxicity. Statistical designs for Phase I/II trials where we combine safety and efficacy information have been investigated ([3, 4]). Phase I trials that include a DEC have a more exploratory flavor to them than Phase I/II designs, which from the outset, balance the two outcomes simultaneously in order to locate the most successful dose. Mostly, before the DEC kicks in, no study of efficacy is made. Those dose finding protocols that involve dose expansion cohorts are still focused on maintaining a safe dose but, simultaneously, keeping a watch over accumulating evidence on efficacy in particular groups. The requirement of the standard Phase I design to treat at the best estimate of the MTD can be relaxed to some degree during the expansion cohort, since there is no longer just one outcome that drives the study.

Our goal for studies that involve a DEC is no longer the same as before where safety alone was evaluated. The current aim of DEC to obtain preliminary efficacy data in a more homogenous patient subgroup(s) (cohorts) is an important distinction. If evidence of efficacy, overall or for some specific groups, is lower than deemed of interest then we may accept an increase in the rate of toxicity in the hope of improving that. Including additional patients during the expansion phase enables us to better estimate the rate of toxicity and the rate of efficacy and provides an indication as to whether we may wish to increase or decrease a dose suggested as the RP2D.

2 Dose allocation during the expansion cohort

We suggest that a model based design be used to guide dose allocation during the DEC whether or not the first part of the study used a model based design. The Continual Reassessment Method (CRM) [10] is a natural candidate, although other model based designs such as escalation with overdose control are good alternatives. The trial consists of k ordered dose levels, $d_1, d_2, \ldots, d_k$, and a total of N patients. The dose level for patient j is denoted as X_j, and the binary outcome DLT yes or no, is denoted as Y_j, where a 1 indicates a DLT. At dose $X_j = x_j$ the probability that $Y_j = 1$ is given by:

$$R(x_j) = Pr(Y_j = 1 | X_j = x_j), \quad x_j \in \{d_1, d_2, \ldots, d_k\}.$$

The estimated MTD is defined to be the dose $d_m \in \{d_1, \ldots, d_k\}$, $1 \le m \le k$ such that, $d_m = \arg\min_{di} \Delta(\hat{R}(d_i), \theta)$, $i = 1, \ldots, k$. where $\Delta(\hat{R}(d_i), \theta)$ indicates the distance from the target acceptable rate θ. The working model is chosen as, $\psi(d_i, a) = \alpha_i^a$, where $a \in (0, \infty)$ is the unknown parameter, and α_i are the standardized units, or skeleton, representing a transformation of the doses at the discrete dose levels d_i. The parameter estimate $\hat{a}$ can be obtained through a Bayesian framework or maximum likelihood estimation (MLE)[11]. The score equation obtained by calculating the slope of the log-likelihood, after the inclusion of the first j patients, can be written:

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

and where $n_i(j)$, $t_i(j)$ denotes the number of patients treated and the number of DLTs observed at level i respectively. This equation can then be solved to obtain the running estimate $\hat{a}$ which is then plugged in to $\hat{R}(d_i) = \phi(d_i, \hat{a})$.

Further to monitoring the rates of toxicity on the patients of the DEC we can also collect information on other indicators of efficacy, or other endpoints such as pharmacokinetic/pharmacodynamic endpoints [13, 12]. We denote the probability of efficacy response at $X_j = x_j$ by: $Q(x_j) = \Pr(V_j = 1 | X_j = x_j)$, where V_j is a random variable taking the value 1 in the presence of a positive outcome, and is zero otherwise. A working model can be used to characterize this, i.e., $\phi(d_i, b) = \beta_i^b$ for $Q(x_j)$ [3], where β_i is some chosen skeleton of initial probabilities of efficacy. Assuming conditional independence between safety and efficacy given the dose, the respective contributions to the likelihood are orthogonal. Letting $n_{e,i}(J)$ be the number of patients treated at dose i who are also evaluable for efficacy response and $r_i(J)$ the number of responders observed at dose i then the estimating equations are then:

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

where $\mathscr{F}_j = \{(x_1, y_1, v_1), \ldots, (x_j, y_j, v_j)\}$. All N patients provide information on safety measurements, in addition to which the J, $J \leq N$ included in the DEC also have information on efficacy.

3 Selection of the recommended Phase II dose

Our purpose is to identify a dose that we would like to take forward into Phase II testing and for this we need to see evidence of efficacy in a disease-specific or histology-specific cohorts. This can be made more formal in a testing framework. Suppose that we have in mind some low rate of efficacy, say q_0, and some higher, more effective rate, q_1, $0 < q_0 < q_1 < 1$. The hypotheses: $H_0 : Q(d_i) \leq q_0$ can then be tested against $H_1 : Q(d_i) \geq q_1$, where $Q(d_i)$ denotes the true, unknown, efficacy rate at level d_i. The safety criterion , i.e. the definition of d_m, determines the dose allocation and ensures we are experimenting at the level that is closest to the target toxicity rate. Given the level is safe, the RP2D at the end of the study is the dose that is determined to be efficacious by one of the test statistics that we present below (T_1, T_2, T_3). The choice of which statistic to use depends on the clinical setting. The test statistics can be based on model based rates (T_1) or on the empirical efficacy rates (T_3) or on the combination of both efficacy and safety data (T_2).

Specifically, at level d_i, the hypotheses of non efficacious versus efficacious drug above 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. A smaller b corresponds to a higher response rate. These are interval rather than point hypotheses and so we need to consider the regions corresponding to either hypothesis. Under H_0 we define the region B_0 as (b_0, ∞) and similarly under H_1 the region B_1 is $(0, b_1)$. For given data $\mathscr{F}_{j^*}$, letting j^* indicate the number of patients treated at level d_i, and the prior for b denoted by $g(b)$, the test statistic can be expressed as:

$$T_1(d_i) = \frac{\int_{B_1} \prod_{l=1}^{j^*} \beta_i^{bv_l} \left(1-\beta_i^b\right)^{(1-v_l)} g(b)\, db}{\int_{B_0} \prod_{l=1}^{j^*} \beta_i^{bv_l} \left(1-\beta_i^b\right)^{(1-v_l)} g(b)\, db} \quad (3)$$

When $T_1(d_i) > (1-\varepsilon_2)/\varepsilon_1$, the test provides support in favor of H_1 and the boundaries can be obtained as a result of fixing the error rates, ε_1, and ε_2 [14, 15]. In common to many sequential procedures, we can fail to make a conclusive decision either way in which case we can view the conclusion to be that for further experimentation. If the test provides enough evidence in favor of H_0, then there would be little enthusiasm for taking the drug or combination at this level forward for further testing. We can test $H_0 : Q(d_i) \leq q_0$ against $H_1 : Q(d_i) \geq q_1$ while controlling for the rate of toxicity via; $H_0 : R(d_i) > s_0$ against $H_1 : R(d_i) \leq s_1$, $0 < s_1 \leq s_0 < 1$. We can denote the restricted spaces A_1 and A_0 for the parameter a under H_1, H_0, and B_1, B_0 denote the space for b under H_1, H_0. Then, given the observations and the prior distributions, $g_1(a)$ and $g_2(b)$, we define:

$$\mathscr{H}_{\{A,B\}} = \int_B \int_A \prod_{l=1}^{j^*} \beta_i^{bv_l} \left(1-\beta_i^b\right)^{(1-v_l)} \prod_{l=1}^{j^*} \alpha_i^{ay_l} \left(1-\alpha_i^a\right)^{(1-y_l)} g_1(a)\, g_2(b)\, da\, db \quad (4)$$

We construct a test on the basis of the ratio of $\mathscr{H}$ integrated over the respective regions under H_1 and H_0. This gives us as a test, $T(d_i) = \mathscr{H}_{\{A_1,B_1\}} / \sum_{\mathscr{R}} \mathscr{H}_{\{A_{\mathscr{R}},B_{\mathscr{R}}\}}$ where $\mathscr{R}$ indicates the relevant regions of interest for the two parameters. Under the null, these can be made to correspond to different hypotheses. If we wish to test for evidence of efficacious and non toxic dose against non-efficacious and toxic dose, then we make use of; $H_0 : b \geq b_0$ and $a \leq a_0$ against $H_1 : b \leq b_1$ and $a \geq a_1$ and the test statistic equals to $\left[\mathscr{H}_{\{A_0,B_0\}}\right]^{-1} \mathscr{H}_{\{A_1,B_1\}}$. If we are at level d_i, $A_0 : (0, a_0)$, $A_1 : (a_1, \infty)$ and $B_0 : (b_0, \infty)$ and $B_1 : (0, b_1)$ are the corresponding regions for the hypothesis given above then:

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

A very useful approximation to composite tests based on interval hypotheses is to find two simple point hypotheses that lead to comparable operating characteristics. [3]

Suppose that patient j is being treated at dose d_i and we wish to test the hypotheses: $H_0 : Q(d_i) = q_0$ against $H_1 : Q(d_i) = q_1$, where q_0 indicates too low a rate and q_1 a satisfactory rate. When $r_i(j)$ is the sum of responders treated at d_i who also have efficacy response measured, we have;

$$T_3(d_i) = r_i(j)\ log\left(\frac{q_1(1-q_0)}{q_0(1-q_1)}\right) + j\ log\frac{(1-q_1)}{(1-q_0)} \quad (5)$$

The idea here is to guide decision making in formal terms for choosing a level to be taken forward into Phase II studies.

4 Simulation study

4.1 Operating Characteristics

In the simulation study we evaluated two dose escalation approaches: 1) investigators follow the 3+3 design [16] during the dose escalation phase, as this is typical in Phase I trials in oncology, followed by a DEC of additional J patients which is guided by a model (CRM model);(scheme A); 2) both the dose escalation and dose expansion phases are guided by the model following the completion of a smaller Phase I trial (scheme B). We assume that the trial established an initial estimate of the MTD using the dose-escalation design, denoted as 3+3 MTD or d_m depending on the scheme above. Additional data, such as efficacy or PK response are obtained during the expansion phase. During the dose expansion, dose allocation is guided by safety, i.e. efficacy is not taken into account in terms of allocating patients to dose levels. In addition allocation during the DEC does not have to be at a single level. For scheme A), patients in DEC are randomized 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 $\hat{R}(d_m) > \theta$. The distance that either level is away from the θ is used as a basis for a biased randomization scheme. Scheme B is allocating patients at the current estimate d_m without randomization. We calculated the sequential tests after each patient, using different forms of hypothesis testing as shown in Section 3, in trials obtained under scheme B. 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 1. 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 $s_1 = s_0 = \theta = 30\%$. 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 for the method since the model here was generated by the true rates.
2. The sample size during the 3+3 stage varies since the trial stops at any time after observing 2 or more DLTs/6 patients 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.

3. There is an MTD defined by the 3+3 during the dose escalation and this is denoted as 3+3 MTD or pre-expansion MTD. Using the dose expansion data, there is also a model-based MTD, d_m, defined based on the predicted toxicity rates and the distance to the target rate, θ. When both the dose escalation and dose expansion phases are guided by the model, then the MTD is defined as d_m.

4. The skeleton values for toxicity and efficacy for the parameters α_j, β_l are given as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 for the 6 levels respectively.

5. For sequential probability ratio tests (SPRT), Type I and Type II errors, ε_1, ε_2 are set at 10%. The uniform distribution was used for the distribution of $g_1(a)$ and $g_2(b)$ in the test statistics T_1, T_2.

4.2 Results

Table 2 provides a summary of the dose recommendation across many simulated trials at the end of a 3+3 design and at the end of the dose expansion phase. 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 and all scenarios showed improvement compared to the MTD found during the dose escalation phase alone. The increase in accuracy is apparent across many scenarios regardless of the location of the MTD. As pointed out by a reviewer, the amount of improvement also depends on the target rate θ we use during the model based/expansion phase (Table 3). Since the 3+3 underestimates the MTD compared to model-based designs and it often selects levels with rates around 0.15-0.25 [16] then a lower θ will be more 'in agreement' with the levels selected pre-expansion. The results indicate that by taking into account the toxicity responses from the additional patients accrued during the expansion phase, we can improve upon our estimated MTD which agrees with previous reports [1]. The efficacy responses are considered secondary in terms of the DEC aims thus are not being used in deciding the dose allocation. The improvements in accuracy are even greater if we use a model- based design from the beginning of the trial and of course greater improvements come with larger sample size (Tables 4, 5). The choice of θ does not affect improvement in this setting, as much as the steepness of the dose-toxicity curve and how separated the levels are in terms of DLT rates.

The use of randomization in the DEC and the use of a model following a 3+3 design helps to move experimentation closer to the true MTD. There are many ways that we can experiment at more than just a single level, for example here we randomized patients to levels, d_m and the level just above d_m if $\hat{R}(d_m) < \theta$ or the level just below d_m if $\hat{R}(d_m) > \theta$. The distance that either level is away from θ, the target rate, was used as a basis for a biased randomization scheme or we could use no more than the fact that one level is believed to be below the MTD and the other above.[17]. We may decide to randomize, on average, twice as many patients below as above, or, more likely, a simple coin toss providing on average the same number of patients allocated above as below.

In order to assess the performance of the SPRT we calculated the proportion of trials reaching a decision in favor of each hypothesis at the MTD and nearby levels, and the average sample number. Table 6 shows the results from carrying out sequential probability ratio tests (SPRT) when the number of patients at the DEC is fixed at $J = 12, 25, 50$. For each one of the three tests we report the percent of trials that reached a decision in favor of each hypothesis at the MTD after a fixed number of patients J has been accrued. 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. All tests correctly accept MTD+1 as an efficacious level, but fail to support MTD as efficacious level, given the fact that the majority of the trials decided in favor of H_0. Note that T_3 supports more frequently continuation of the trial given the small sample size and lack of efficacy responses in the empirical rates. However, T_2 tests simultaneously for a safe and efficacious level, and is model-based tests thus they reach a conclusion in favor of H_1 or H_0 more often compared to T_3. In Scenario 2, the tests correctly identify the MTD as efficacious and safe dose which is supported by the true rates. 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 for the MTD. The last scenario represents a case where the working models follow the true rates and the MTD is efficacious. In this scenario, 65%-83% of trials reached a decision in favor of H_1 or H_0 after 25 patients. Note that 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 the estimated Type I error is less than the set value of 10% (Table 6). In Scenario 2 the MTD has an efficacy rate of 30%, thus 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% in almost all cases.

In practice, it is useful to know how many patients are required to make a decision. The value of average sample number (ASN) indicates when the test statistic crosses one of the two boundaries and the sequential procedure terminates. For these calculations we truncated the SPRT calculations at a maximum sample of 200 patients. Figure 1 shows the ASN for each test statistic. The median sample size is smaller with T_2 compared to the test based on the empirical rates(T_3). This indicates that we can obtain increase in efficiency by using the modeled complete data on toxicity and efficacy rather than the simple empirical rates. Using a model based test, we need on average 10-20 patients to reach a decision which is consistent with current, clinical practice.

4.3 Sensitivity analysis

In the above simulations a Uniform distribution was used for the parameters a, b, according to previous findings that supported the use of a non informative distribution such as Uniform. A sensitivity analysis for the choice of distribution $g(b)$ was carried out and reported in a previous report [15]. In the 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 showed that that in certain cases, a Gamma distribution with small variance can be more informative than a Uniform distribution as the sequential test was reaching a decision early on and a higher number of trials were reaching a decision in favor of H_1 [15]. The relative importance of the prior depends on the sample size and the influence of the prior matters early on in the trial. Here we are using sequential

tests to guide a decision process of whether to terminate or continue the trial based on the number of efficacy responses and simultaneously we are estimating the location of the MTD. If the decision to stop the trial early is influenced by the prior then we will never collect more data to allow overriding the prior. Finally, the robustness of our proposed approach to the assumption of conditional independence under various values of correlation parameter was assessed by additional simulations and our results supported that ignoring the dependence structure, even when strong, has a negligible impact in terms of the decisions reached by the proposed tests, or model parameter estimation, which agrees with recent findings [18, 15]. Although it may be possible to work with models that include terms quantifying the degree of correlation between the efficacy and toxicity outcomes, this is not usually helpful. All of the information that we need is contained in the marginal rates of efficacy and toxicity. Both of these are driven by the dose and it is certainly a reasonable working assumption that, given dose, the two related outcomes can be treated as though they were independent.

Finally, our assumption is that there is homogeneity in terms of safety, and thus the data from patients accrued before or after expansion can be combined in a single likelihood (or that the toxicity data from the phase I part are exchangeable with those from the cohort expansion). However, heterogeneity is likely to exist in terms of efficacy if we accrue multiple cohorts with specific patient characteristics in each cohort. This paper focused on a cohort specific analysis.

5 Discussion

The trend toward the design and analysis of Phase I studies based on some kind of model has grown steadily in the past twenty or so years. The standard 3+3 design cannot be used to address the questions raised by DEC. With recent developments whereby the addition of a dose expansion cohort has become almost routine we need give more thought to both the design and analysis of such studies. The methodology described here provides more support to designing and monitoring expansion cohorts in Phase I oncology trials. Until very recently the idea of adding DEC was viewed as an add on (ad-hoc) to the dose escalation design and often the enrollment of a fixed number of patients treated at a single level was deemed sufficient. Currently, the inclusion of DEC is becoming routine in multicenter Phase I trials and the clinical aims of DEC are evolving. Given our evaluation here and in other reports ([15]) we believe the use of a model based SPRT based on efficacy (T_1) addresses the aims of DEC while it is simpler than T_2 and more efficient than T_3. In practice though, the choice of which statistic to use depends on whether we want to test for both safety and efficacy (T_2) or efficacy alone (T_1).

In common with other sequential adaptive procedures the aim is to make a sharp choice between concluding in favor of the dose level, rejecting the dose level due to insufficient activity or to a decision of inconclusiveness whereby we can stop and declare the question unanswered or we can continue until able to make a more clear cut decision. Such decision making has very much the flavor of the decision-making process of a Phase II. The ethical restrictions of a pure Phase I study indicate that patients be treated as close as we can to the current estimate of the MTD. However, once we embark on an expansion cohort and we are

no longer looking at just evidence of tolerability then it is no longer at all clear, nor easy to justify, that experimentation ought be concentrated at a single level. A higher level may provide a much better response rate with little increase in toxicity; a lower level may provide a similar efficacy rate with a noticeable reduction in toxicity. In this sense, the inclusion of an expansion cohort changes the rationale and, indeed, the goals themselves of the study.

Whether or not a Phase I dose finding study that includes a DEC can be viewed as a seamless Phase I/Phase II study, there will certainly be cases in which enough evidence has been gathered in the DEC to move very quickly to a large scale randomized clinical trial. It would be premature to argue that the Phase II study can be supplanted by a more sophisticated Phase I study plus a DEC but, in terms of gathering all the evidence we can before embarking on the large scale Phase III study, it is clear that the DEC provides an additional tool. Part of the DEC will say something about the potential for treatment effect and, simultaneously, a part will help us make any adjustments that are needed to a potentially poor estimate of the MTD from the pre-DEC component of the study as supported by our simulations.

Biomarkers and targeted therapies have had an impact on the design of all phases of clinical trial experimentation. They have no less an impact on the early phase studies and, in particular, those that involve multiple DEC. The structure of the DEC, the number of subgroups to be studied, the relationships between these groups with toxicity and efficacy outcomes, can all depend on specific biomarkers and genetic profiles. A carefully designed Phase I study that includes a DEC has the potential to provide a much more precise setting in which to proceed to a Phase II and Phase III study. In this way, the addition of a DEC to the dose finding study should enable us to reduce the number of Phase II studies that fail to show any effect. It should also help reduce the number of studies that fail at the Phase II stage as a result of the chosen dose level being poorly tolerated even when there is evidence of efficacy.

Acknowledgments

Funding:This work was partially supported by National Institute of Health (Grant Number 1R01CA142859, P30 CA008748).

References

1. Iasonos A, O'Quigley J. Design considerations for dose-expansion cohorts in phase I trials. Journal of Clinical Oncology. Nov 1; 2013 31(31):4014–21. [PubMed: 24101039]
2. Iasonos A, O'Quigley J. Clinical trials: Early phase clinical trials-are dose expansion cohorts needed? Nat Rev Clin Oncol. Oct 6.2015 doi: 10.1038/nrclinonc.2015.174.
3. O'Quigley J, Hughes MD, Fenton T. Dose-finding designs for HIV studies. Bio-metrics. Dec; 2001 57(4):1018–29.
4. Yuan Y. Bayesian Phase I/II adaptively randomized oncology trials with combined drugs. The Annals of Applied Statistics. 2011; 5:924–942. [PubMed: 22375162]
5. Ivanova A. A New Dose-Finding Design for Bivariate Outcomes. Biometrics. Dec; 2003 59(4): 1001–1007. [PubMed: 14969479]
6. Rogatko A, Schoeneck D, Jonas W, et al. Translation of innovative designs into phase I trials. J Clin Oncol. 2007; 25(31):4982–6. [PubMed: 17971597]

7. 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. [PubMed: 22658127]
8. 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. Nov 20; 2013 31(33):4260–7. [PubMed: 24127441]
9. 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. [PubMed: 22261804]
10. 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. [PubMed: 2350571]
11. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. Biometrics. 1996; 52:673–84. [PubMed: 8672707]
12. O'Quigley J, Hughes MD, Fenton T, Pei L. Dynamic calibration of pharmacokinetic parameters in dose-finding studies. Biostatistics. 2010; 11(3):537–545. [PubMed: 20203258]
13. Pei L, 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; 64(4):11171125.
14. Wald, A. Sequential Analysis. John Wiley and Sons; New York: 1947.
15. Iasonos A, O'Quigley J. Dose expansion cohorts in Phase I trials. Statistics in Biopharmaceutical Research. 2016 in press.
16. 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. [PubMed: 18827039]
17. O'Quigley J, Conaway M. Continual Reassessment and related dose finding designs. Statistical Science. 2010; 25(2):202–216. [PubMed: 21731191]
18. Cunanan K, Koopmeiners JS. Evaluating the Performance of Copula Models in Phase I-II Clinical Trials Under Model Misspecification. BMC Medical Research Methodology. 14 Article 51, 2014.

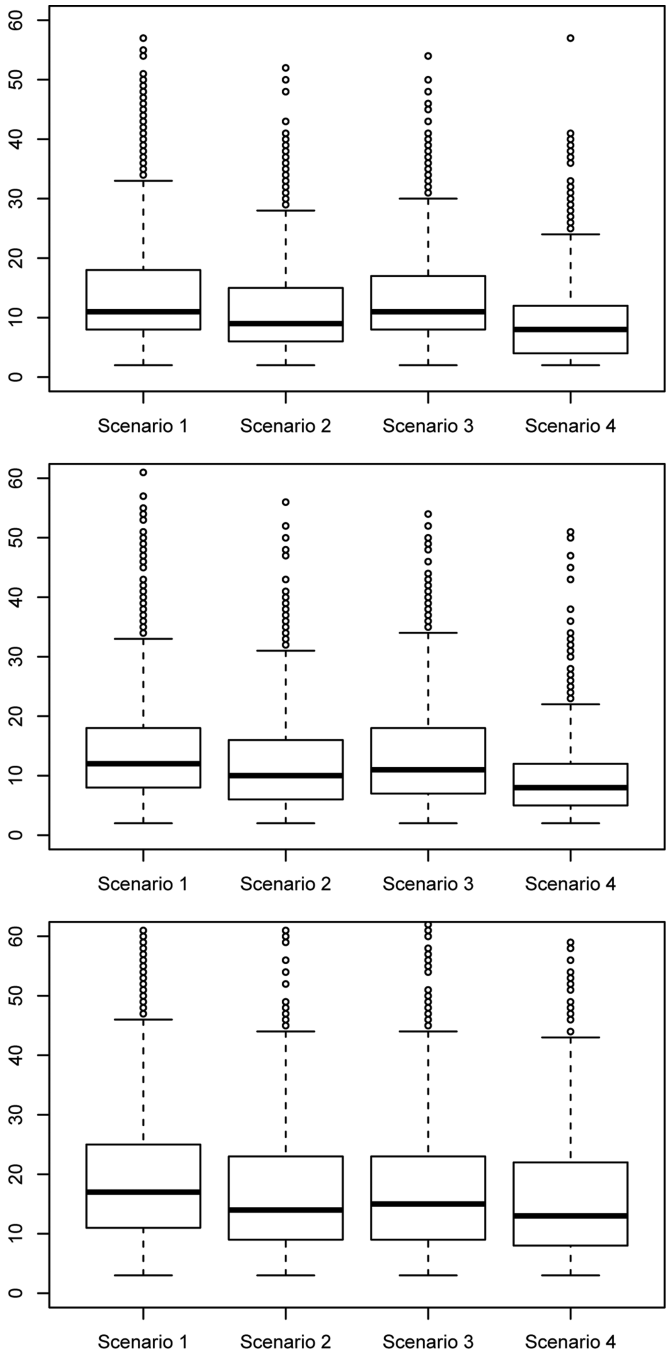

Figure 1.

Distribution of sample size required to make a decision in favor of H_1 or H_0 at any dose level. Horizontal panels show sequential tests based on T_1(top), T_2(middle), T_3 (bottom) for the 4 scenarios respectively.

Table 1

True toxicity and efficacy rates used in simulation study, denoted by $R(d_i)$, $Q(d_i)$ respectively.

Dose Levels	**1**	**2**	**3**	**4**	**5**	**6**
Scenario 1						
$R(d_i)$	0.10	0.15	0.30	0.45	0.50	0.60
$Q(d_i)$	0.05	0.09	0.10	0.30	0.40	0.45
Scenario 2						
$R(d_i)$	0.05	0.10	0.15	0.20	0.30	0.60
$Q(d_i)$	0.01	0.05	0.09	0.10	0.30	0.40
Scenario 3						
$R(d_i)$	0.05	0.10	0.15	0.27	0.33	0.60
$Q(d_i)$	0.01	0.05	0.09	0.10	0.30	0.40
Scenario 4						
$R(d_i)$	0.02	0.06	0.12	0.20	0.30	0.41
$Q(d_i)$	0.02	0.06	0.12	0.20	0.30	0.41

Table 2

Percent of trials selecting each dose. Trials followed the 3+3 followed by a model-based dose escalation during the expansion phase (with randomization proportional to the distance of the estimated rates to the target rate, $\theta = 0.3$.). Median of $\hat{a}$ across simulations is given after J patients are included in the DEC.

Dose Levels	**0**	**1**	**2**	**3**	**4**	**5**	**6**	$\hat{a}$
Scenario 1		0.10	0.15	**0.30**	0.45	0.50	0.60	
3+3 (preexpansion)	0.08	0.19	0.40	**0.26**	0.06	0.01	0.00	
J=12		0.01	0.16	**0.44**	0.26	0.04	0.01	1.04
J=25		0.00	0.14	**0.51**	0.25	0.03	0.00	1.04
J=50		0.00	0.11	**0.61**	0.20	0.01	0.00	1.03
Scenario 2		0.05	0.10	0.15	0.20	**0.30**	0.60	
3+3 (preexpansion)	0.03	0.09	0.17	0.23	0.26	**0.20**	0.01	
J=12		0.00	0.02	0.10	0.28	**0.45**	0.12	1.61
J=25		0.00	0.00	0.06	0.27	**0.54**	0.10	1.64
J=50				0.02	0.26	**0.63**	0.07	1.66
Scenario 3		0.05	0.10	0.15	0.27	0.33	0.60	
3+3 (preexpansion)	0.02	0.10	0.17	0.36	**0.20**	**0.13**	0.02	
J=12		0.00	0.02	0.12	**0.38**	**0.38**	0.07	1.48
J=25		0.00	0.01	0.10	**0.40**	**0.41**	0.06	1.49
J=50				0.05	**0.45**	**0.43**	0.04	1.50
Scenario 4		0.02	0.06	0.12	0.20	**0.30**	0.41	
3+3 (preexpansion)		0.04	0.12	0.27	0.28	**0.20**	0.08	
J=12		0.00	0.00	0.04	0.24	**0.48**	0.23	1.75
J=25		0.00	0.00	0.02	0.20	**0.56**	0.22	1.76
J=50		0.00	0.00	0.01	0.18	**0.63**	0.19	1.75

Table 3

Percent of trials selecting each dose. Trials followed the 3+3 followed by a model-based dose escalation during the expansion phase (with randomization proportional to the distance of the estimated rates to the target rate, θ = 0.2.). Median of $\hat{a}$ across simulations is given after J patients are included in the DEC.

Dose Levels	**0**	**1**	**2**	**3**	**4**	**5**	**6**	$\hat{a}$
Scenario 1		0.10	**0.15**	0.30	0.45	0.50	0.60	
3+3 (preexpansion)	0.08	0.19	**0.40**	0.27	0.06	0.01	0.00	
J=12		0.11	**0.43**	0.34	0.04	0.00	0.00	1.10
J=25		0.09	**0.50**	0.30	0.03	0.00	0.00	1.10
J=50		0.06	**0.55**	0.31	0.01	0.00	0.00	1.10
Scenario 2		0.05	0.10	0.15	**0.20**	0.30	0.60	
3+3 (preexpansion)	0.03	0.09	0.17	0.23	**0.26**	0.20	0.01	
J=12		0.02	0.12	0.29	**0.35**	0.17	0.02	1.60
J=25		0.01	0.08	0.31	**0.41**	0.15	0.01	1.62
J=50			0.06	0.30	**0.47**	0.14	0.00	1.65
Scenario 3		0.05	0.10	**0.15**	0.27	0.33	0.60	
3+3 (preexpansion)	0.02	0.10	0.17	**0.36**	0.20	0.13	0.02	
J=12		0.03	0.13	**0.37**	0.36	0.09	0.00	1.48
J=25		0.01	0.12	**0.42**	0.36	0.07	0.001	1.49
J=50			0.07	**0.49**	0.38	0.03	0.00	1.49
Scenario 4		0.02	0.06	0.12	**0.20**	0.30	0.41	
3+3 (preexpansion)	0.00	0.04	0.12	0.27	**0.28**	0.20	0.08	
J=12		0.00	0.04	0.25	**0.43**	0.21	0.05	1.74
J=25		0.00	0.04	0.24	**0.47**	0.24	0.02	1.75
J=50		0.00	0.00	0.23	**0.55**	0.21	0.01	1.76

Table 4

Percent of trials selecting each dose. Trials followed a model-based dose escalation design (CRM, with target rate $\theta = 0.3$) from the beginning including the expansion phase (without randomization). Median of $\hat{a}$ across simulations is given after J patients are included in the DEC.

Dose Levels	**1**	**2**	**3**	**4**	**5**	**6**	$\hat{a}$
Scenario 1	0.10	0.15	**0.30**	0.45	0.50	0.60	
CRM (preexpansion)	0.02	0.19	**0.44**	0.24	0.07	0.02	1.04
J=12	0.01	0.19	**0.55**	0.22	0.03	0.00	1.01
J=25	0.00	0.16	**0.64**	0.19	0.02	0.00	1.01
J=50		0.08	**0.79**	0.13	0.00	0.00	1.00
Scenario 2	0.05	0.10	0.15	0.20	**0.30**	0.60	
CRM (preexpansion)		0.02	0.11	0.31	**0.46**	0.10	1.57
J=12		0.01	0.05	0.31	**0.56**	0.07	1.61
J=25		0.00	0.06	0.28	**0.63**	0.04	1.63
J=50			0.01	0.24	**0.71**	0.04	1.66
Scenario 3	0.05	0.10	0.15	**0.27**	**0.33**	0.60	
CRM (preexpansion)		0.02	0.17	**0.38**	**0.36**	0.06	1.45
J=12		0.01	0.10	**0.45**	**0.39**	0.05	1.45
J=25		0.00	0.10	**0.47**	**0.41**	0.02	1.46
J=50			0.04	**0.50**	**0.44**	0.01	1.48
Scenario 4	0.02	0.06	0.12	0.20	**0.30**	0.41	
CRM (preexpansion)		0.00	0.05	0.27	**0.43**	0.24	1.69
J=12		0.00	0.02	0.26	**0.52**	0.20	1.71
J=25			0.01	0.24	**0.58**	0.18	1.71
J=50			0.00	0.18	**0.66**	0.15	1.71

Table 5

Percent of trials selecting each dose. Trials followed a model-based dose escalation design (CRM, with target rate $\theta = 0.2$) from the beginning including the expansion phase (without randomization). Median of $\hat{a}$ across simulations is given after J patients are included in the DEC.

Dose Levels	**1**	**2**	**3**	**4**	**5**	**6**	$\hat{a}$
Scenario 1	0.10	**0.15**	0.30	0.45	0.50	0.60	
CRM (preexpansion)	0.19	**0.43**	0.29	0.09	0.01	0.00	1.10
J=12	0.14	**0.49**	0.34	0.03	0.00	0.00	1.10
J=25	0.10	**0.55**	0.34	0.01	0.00	0.00	1.11
J=50	0.05	**0.65**	0.29	0.00	0.00	0.00	1.10
Scenario 2	0.05	0.10	0.15	**0.20**	0.30	0.60	
CRM (preexpansion)	0.03	0.14	0.28	**0.35**	0.19	0.01	1.61
J=12	0.01	0.10	0.32	**0.41**	0.16	0.01	1.62
J=25	0.00	0.09	0.35	**0.42**	0.14	0.00	1.61
J=50	0.00	0.04	0.35	**0.51**	0.10	0.00	1.66
Scenario 3	0.05	0.10	**0.15**	0.27	0.33	0.60	
CRM (preexpansion)	0.03	0.16	**0.38**	0.30	0.12	0.00	1.48
J=12	0.01	0.12	**0.45**	0.35	0.07	0.00	1.49
J=25	0.00	0.11	**0.49**	0.35	0.05	0.00	1.49
J=50	0.00	0.05	**0.55**	0.38	0.02	0.00	1.49
Scenario 4	0.02	0.06	0.12	**0.20**	0.30	0.41	
CRM (preexpansion)	0.00	0.05	0.30	**0.39**	0.21	0.04	1.69
J=12	0.00	0.03	0.29	**0.48**	0.18	0.01	1.71
J=25	0.00	0.02	0.28	**0.54**	0.16	0.00	1.70
J=50	0.00	0.00	0.22	**0.66**	0.13	0.00	1.73

Table 6

Proportion of trials deciding in favor of H_1, H_0, or to continue the trial (inconclusive) when including J = 12,25,50 patients treated at the DEC. The four scenarios represent trials simulated and presented in Table 4 (CRM trials with $\theta = 0.3$).

		Scenario1			Scenario 2		
	J	***H*** **1**	***H*** **0**	**continue**	***H*** **1**	***H*** **0**	**continue**
T_1	50	0.03	0.90	0.08	0.77	0.10	0.13
	25	0.03	0.70	0.27	0.56	0.09	0.35
	12	0.05	0.49	0.46	0.44	0.10	0.46
T_2	50	0.04	0.85	0.12	0.78	0.09	0.13
	25	0.06	0.69	0.24	0.63	0.11	0.25
	12	0.10	0.46	0.44	0.47	0.14	0.39
T_3	50	0.03	0.81	0.16	0.79	0.08	0.12
	25	0.02	0.70	0.28	0.55	0.08	0.37
	12	0.03	0.28	0.69	0.42	0.04	0.55
	J	**Scenario 3**			**Scenario 4**		
T_1	50	0.37	0.54	0.09	0.83	0.03	0.14
	25	0.29	0.43	0.27	0.66	0.04	0.30
	12	0.26	0.29	0.45	0.54	0.08	0.39
T_2	50	0.38	0.51	0.11	0.86	0.03	0.11
	25	0.34	0.42	0.24	0.72	0.06	0.22
	12	0.29	0.29	0.42	0.55	0.10	0.35
T_3	50	0.39	0.49	0.12	0.86	0.01	0.13
	25	0.28	0.44	0.29	0.65	0.04	0.32
	12	0.21	0.15	0.64	0.51	0.02	0.47