

# Phase I Clinical Trial Design

Lawrence V Rubinstein, PhD\*

Richard M Simon, DSc\*

Biometric Research Branch, National Cancer Institute

6130 Executive Blvd, Suite 8130, MSC 7434

Bethesda, MD 20892-7434

Phone: 301-496-4836

FAX: 301-402-0560

[rubinsteinl@ctep.nci.nih.gov](mailto:rubinsteinl@ctep.nci.nih.gov)

[rsimon@nih.gov](mailto:rsimon@nih.gov)

( To Appear in Handbook of Anticancer Drug Development)

---

\* Authors contributed equally and are listed alphabetically

## Introduction

The objective of a phase I trial is to determine the appropriate dosage of an agent or combination to be taken into further study and to provide initial pharmacologic and pharmacokinetic studies. It is generally assumed, at this stage of testing, that increased dose is associated with increased chance of clinical efficacy. Therefore, the phase I trial is designed as a dose-escalation study to determine the maximum tolerable dosage (MTD), that is, the maximum dose associated with an acceptable level of dose-limiting toxicity (DLT--usually defined to be grade 3 or above toxicity, excepting grade 3 neutropenia unaccompanied by either fever or infection<sup>35</sup>). This MTD is then taken into further testing. Since evaluation of efficacy is generally not the objective of a phase I trial, it is not necessary to restrict to a patient population homogeneous with respect to disease, or even to restrict to patients with measurable disease (for which tumor response is determinable). It is important, however, to exclude patients with impaired organ function, who may therefore be more prone to serious toxicity. The fundamental conflict in phase I trials is between escalating too fast, so as to expose patients to excessive toxicity, and escalating too slow, so as to deny patients the opportunity to be treated at potentially efficacious dose levels<sup>10</sup>. Phase I trials for compounds or biologics in which toxicity is not expected, and determination of the MTD is not the objective, will be discussed later in this chapter.

The first problem in a phase I trial is deciding on a safe, but not overly conservative, initial dose for the trial. If the agent is new to clinical testing, this must be based on animal studies. It has been determined that the dose (defined in mg per meters squared of body surface area) associated with 10% lethality in mice (MELD<sub>10</sub>) can be predicted to be roughly equivalent to the human MTD<sup>18</sup>. This approach is derived from the concept of "allometric scaling"<sup>15,25</sup>. Toxicity as a function of body weight or surface area is assumed to be roughly constant across species. The initial dose for the phase I trial is taken to be 1/10 the MELD<sub>10</sub> or, if smaller, 1/3 the LD<sub>10</sub> (associated with 10% lethality) in the beagle dog<sup>23</sup>. The use of a second species has been shown to be necessary, since in approximately 20% of approximately 90 reviewed drugs, mouse data alone was insufficient to safely predict the human MTD<sup>2</sup>. American investigators generally use the dog as the second species, while European investigators generally use the rat, with equivalent safety<sup>2</sup>. The next problem is to define dose increments for the subsequent dose levels, and it is here that the various phase I trial designs part company.

## Standard phase I design

The “standard” phase I design utilizes a set of decreasing “Fibonacci” dose level increments proposed by Schneiderman<sup>32</sup>, and currently taken to be 100%, 67%, 50%, 40%, and 33% thereafter<sup>10</sup>. These increments are added to each dose level to give the succeeding level. In other words, the second dose level is 100% greater than the first, the third is 67% greater than the second, and so forth. The purpose is to allow more aggressive dose escalation for the initial levels, which are expected to be sufficiently removed from the MTD for this to be safe. If the MELD<sub>10</sub> accurately predicted the human MTD, only 5-6 such dose escalations would be necessary to complete a “standard” phase I design. Unfortunately, this is often not the case<sup>27</sup>.

The “standard” rule governing dose escalation from one level to the next relies on no assumptions concerning the shape of the dose-toxicity curve or the potential for cumulative toxicity, and therefore the decision to escalate to the next dose level is based solely on toxicity results from the first course administration of the current level. The dose escalation rules (Table 1) proceed as follows, escalating in cohorts of 3-6 patients per dose level<sup>35</sup>. Three patients are treated at the current dose level. If at least 2 patients are observed to have DLT, the prior dose level is defined as the MTD (unless only 3 patients have been treated at that level, in which case it is the tentative MTD). If 0 of the 3 patients are observed to have DLT, the dose level is escalated one step for the next cohort of 3 patients, and the process continues as above. If exactly 1 of the 3 patients treated show DLT, 3 additional patients are treated at the current dose level. If none of these additional 3 patients show DLT, the dose level is escalated for the next cohort of 3 patients, and the process continues as above; otherwise, the prior dose level is defined as the MTD (unless only 3 patients have been treated at that level, in which case it is the tentative MTD). A tentative MTD becomes final when a total of 6 patients are treated with less than 2 showing DLT.

The statistical operating characteristics of this approach are as follows (Table 2). If at least 2 of 3 patients treated at a particular dose show DLT, we can conclude with 90% confidence that the true probability of DLT at that dose is greater than 20%. (In other words, as we see in Table 2, unless the true probability of DLT at that dose is at least 20%, the probability of at least 2 out of 3 patients exhibiting DLT is less than 10%.) On the other hand, if 0 of 3 patients show DLT, we can conclude with 90% confidence that the true probability of DLT is less than 55%. (Again, as we see in Table 2, unless the true probability of DLT is less than 55%, the probability of 0 out of 3 patients exhibiting DLT is less than 10%.) In the interest of efficiency, we accept either of these situations as

sufficient to halt or continue escalation after treating only 3 patients at the current level. Allowing for expansion to 6 patients in case 1 of the initial 3 show DLT, the dose escalation rule gives 91% probability that dose escalation will not halt at doses associated with DLT probability less than 10%, and it gives 92% probability that escalation will not proceed beyond doses associated with DLT probability in excess of 60% (Table 2). The process of approaching the MTD from below, in successive steps, further protects against defining an MTD associated with excessive toxicity. Table 2 plus simulations<sup>17, 20</sup> show that, for a wide variety of dose-toxicity curves, the probability is approximately 85% - 90% that the defined MTD will be associated with DLT probability of approximately 10% - 45%.

The primary criticisms of the standard phase I design<sup>17, 30, 35, 39</sup> are:

- 1) It does not target a particular probability of DLT to be associated with the MTD, and, in practice, the DLT rate associated with the defined MTD will be somewhat dependent on the DLT rates of the various dose levels.
- 2) The MTD definition is unnecessarily imprecise in that it does not make adequate use of all the available first-course toxicity data.
- 3) The dose escalation is unnecessarily slow, leading to treatment of excessive numbers of patients at dose levels less likely to be efficacious.

Storer<sup>39</sup> proposed defining the MTD by fitting all the first course toxicity data to a logistic dose-toxicity curve (a sigmoidal curve that maps dose levels to associated DLT rates, for example, equation (1), discussed in more detail below) and letting the MTD be the dose level associated with the targeted DLT rate (usually, 20% - 30%), thus addressing criticisms (1) and (2) of the standard design. To address criticism (3), he suggested escalating the dose in single-patient cohorts until DLT is observed, at which point dose escalation would revert to the standard design.

## **Continual Reassessment Method (CRM)**

O'Quigley et al.<sup>30</sup> extended the modeling idea of Storer<sup>39</sup> by proposing the use of a dose-toxicity model to guide the dose-escalation, as well as to define the MTD. First, a statistical model, such as equation (1), relating dose to probability of dose-limiting toxicity, is defined. Using a Bayesian statistical approach<sup>24</sup>, the free parameter ( $\alpha$ )

of the model is initially given a “prior” probability distribution such that the model maps the dose-levels to probabilities of dose-limiting toxicity in accord with investigator expectations. O’Quigley et al.<sup>30</sup> proposed that each successive patient in the phase I trial be treated at the expected MTD, according to the current state of the model, and that the model be immediately “updated” (that the “posterior” distribution of the free parameter be recalculated, according to Bayes’ theorem<sup>24</sup>) by incorporating first-course toxicity data obtained from each successive patient. They proposed that when the sample size reached a preset limit of 20-25, the MTD be calculated from the final state of the dose-toxicity model.

### ***Original form of CRM***

O’Quigley et al.<sup>30</sup> designated the above approach the Continual Reassessment Method (CRM). It can be made clearer by examining use of the following one-parameter logistic model, proposed by Goodman et al.<sup>17</sup> for defining the probability of DLT  $p_i$  at the  $i^{\text{th}}$  dose level, in conjunction with CRM:

$$p_i = e^{3+\alpha \cdot x_i} / (1 + e^{3+\alpha \cdot x_i}) \quad (1)$$

By the methods of Goodman et al.<sup>17</sup> the investigators first define an increasing set of dose levels (indexed by  $i$ ) to be used in the phase I trial. The investigators provide initial expectations of the probabilities of DLT (the  $p_i$ ’s) at those doses. The initial (“prior”) distribution of the parameter  $\alpha$  is taken to be the standard exponential distribution with mean and variance equal to one. The  $x_i$  values are determined by equation (1) by letting  $\alpha$  be equal to one (its mean according to the initially given exponential distribution) and by letting the  $p_i$ ’s be the initial expectations of the investigators. (For example, Goodman et al.<sup>17</sup> give  $x_i$  values of -5.9, -5.2, -4.3, -3.6, -3.0, and -2.15, to correspond to prior expectations for DLT rate  $p_i$  of .05, .1, .2, .35, .5, and .7.) The substantial uncertainty of the investigators’ initial expectations is represented by the variability associated with the initial distribution of  $\alpha$ . For example, using the above prior distribution, the dose initially associated with an expected DLT rate of 20% has a 33% probability of actually being associated with a DLT rate in excess of 75%, and it has a 20% probability of being associated with a DLT rate less than 5%. (In other words, the initial state of the model reflects that the

investigators' initial guess at an MTD could actually be either a very toxic dose, or a very non-toxic dose, both with reasonably high probability.) As each successive patient is treated, the distribution of  $\alpha$  is re-calculated according to Bayes' theorem<sup>24</sup>, to reflect the new toxicity data and the greater certainty associated with the dose-toxicity relationship. Equation (1), with  $\alpha$  having this re-calculated "posterior" distribution, eventually reflects the dose-toxicity pattern actually observed in the phase I trial, with substantially less uncertainty associated with the predicted DLT rates  $p_i$ .

O'Quigley et al.<sup>29,30</sup> suggested fixing the sample size of a CRM-based phase I trial at 20-25 patients. At the termination of the trial, the MTD is defined to be the dose associated with the target DLT rate (usually 15% - 25%), according to the final state of the dose-toxicity model (according to equation (1), for example, letting  $\alpha$  be the mean of its final "posterior" distribution). O'Quigley<sup>28</sup> gives simulations to demonstrate the accuracy of the confidence interval for the rate of DLT at the chosen MTD (for sample size 20). O'Quigley et al.<sup>29,30</sup> argued that CRM addresses the serious concerns associated with the standard phase I design, given above. They noted that use of a dose-toxicity model allows the investigators to target a specific DLT rate to be associated with the MTD, and it allows all of the first-course toxicity to be incorporated in defining the MTD. They stressed the importance of treating each patient at a sufficiently high dose to offer the hope of an effect, and they asserted that treating each patient at the currently estimated MTD avoids systematic under-treatment of patients, without involving significantly increased risk of DLT (compared to the standard design), according to their simulations<sup>29</sup>.

### ***Amendments and alterations of CRM***

Korn et al.<sup>20</sup> argued that, based on their simulations, CRM did, in fact, significantly increase the DLT risk to patients, compared to the standard design. They demonstrated that CRM tended, with substantially increased probability, to treat patients at doses higher than the MTD, even at doses two or more levels higher, where DLT could be not only more frequent, but also more serious. This was seen to be a result of treating each successive patient at the currently estimated MTD. In particular, the initial patients were to be treated thus, despite the fact that the initial state of the dose-toxicity model might often reflect the uncertainty of the investigators with respect to the clinical toxicity of the untested agent.

Concerns such as these resulted in a number of proposed alterations to the original CRM. Goodman et al.<sup>17</sup> suggested that dose escalation begin at the standard initial dose (usually the MELD<sub>10</sub>), and that it proceed, at most, one dose step at a time (although they did not give guidance as to how these dose steps should be defined). They presented simulations to demonstrate that this approach avoided the increased DLT risk associated with the original CRM, while preserving the advantages of greater efficiency and accuracy. Babb et al.<sup>3</sup> suggested that, rather than treat patients at the dose expected to yield the targeted rate of DLT (which gives 50% likelihood of exceeding the targeted MTD, according to the dose-toxicity model), patients should be treated at the dose associated with 25% likelihood of exceeding the MTD, according to the current state of the model. They presented simulations to demonstrate that this approach also avoided the increased DLT risk of the original CRM, while preserving efficiency and accuracy. Finally, Potter<sup>31</sup> suggested, in answer to concerns about attempting to define an initial dose-toxicity model without clinical experience, that the initial stage of the phase I trial proceed in a standard fashion (escalating from the starting dose with successive 50% dose increments), until DLT is observed. At that point, a dose-toxicity model would be constructed, based on the trial data only. Patients would then be treated at the currently estimated MTD, based on the model. The trial would terminate when 18 patients had been treated, with at least 4 instances of DLT, and with at least 9 patients treated subsequent to the initial such instance.

All of the above alterations were accompanied by a retreat from single-patient cohorts, as originally suggested by O'Quigley et al.<sup>29, 30</sup>, to three-patient cohorts. This was prompted, in part, by the practical consideration relating to the usual brisk accrual to phase I trials, but also, more importantly, by the desire to achieve greater safety with the accumulation of more first-course toxicity data between successive updates of the dose-toxicity model.

## **Accelerated Titration Designs**

Accelerated titration designs attempt to improve several aspects of conventional designs. (i) With standard designs many patients are treated at doses well below the biologically active level, minimizing the opportunity for anti-tumor response. (ii) Many phase I trials using the standard design take a long time to complete. (iii) Conventional designs select a dose for the population of patients, and there is no attempt to tailor doses to individual patients; (iv) Conventional designs provide little information about inter-patient variability, cumulative toxicity or

the steepness of the dose-toxicity relationships.

Accelerated titration designs are characterized by (i) A rapid initial escalation phase; (ii) Intra-patient dose escalation; and (iii) Analysis of results using a model that incorporates parameters for intra-patient variation in toxic effects, cumulative toxicity and steepness of dose-toxicity effects. The analytic model incorporates data from all courses of therapy and for graded toxicity levels.

### ***Rapid Acceleration Phase***

Simon et al.<sup>35</sup> defined several accelerated titration designs and compared them to a standard design (called design 1). Design 1 differs from the standard phase I design described above only in that Simon et al.<sup>35</sup> used fixed 40% dose steps because it was felt that there is not real justification for the standard Fibonacci approach.

Design 2 utilizes single patient cohorts per dose level during the accelerated phase with 40% dose increments. When the first instance of first-course DLT is observed, or the second instance of first-course intermediate toxicity is observed, the cohort for the current dose level is expanded to three patients and the trial reverts to use of design 1 for further cohorts. “Intermediate toxicity” can be defined in a protocol specific manner. Simon et al.<sup>35</sup> used any grade 2 toxicity that was considered treatment related as intermediate toxicity.

Design 3 is similar to design 2 in that single patient cohorts are used during the accelerated phase. With design 3, however, double dose steps are used during the accelerated phase. Two 40% dose steps corresponds to approximately a doubling of the actual dose. The accelerated phase ends, as with design 2, when the first instance of first-course DLT or the second instance of first-course intermediate toxicity. After that, design 1 is used for all further cohorts.

Design 4 is similar to design 3 in that single patient cohorts and double dose steps are used during the accelerated phase. Design 4 differs from design 3 only in the criterion used for triggering the end of the accelerated phase. With designs 2 and 3, the accelerated phase ends with the first instance of first-course DLT or the second instance of first-course intermediate toxicity. With design 4, the trigger is the first instance of any-course DLT or the second instance of any-course intermediate toxicity. Hence, design 4 may stop the accelerated phase earlier than design 3.



### ***Intra-patient dose escalation***

Accelerated titration designs were designed to permit dose-escalation in subsequent courses for a patient who remains on study and has no evidence of toxicity at the dose used during the current course. The rule used was that if less than intermediate level toxicity is observed for a patient during a course, then the dose is escalated for the next course if that patient stays on study. If intermediate level toxicity occurs, then the dose stays the same for the next course if that patient stays on study. If DLT occurs, then the patient generally goes off study, but if not, then the dose is reduced. For design 2, single dose steps are used for intra-patient dose changes. For designs 3 and 4, double dose steps are used for intra-patient dose changes during the accelerated phase, and single dose steps subsequently. The accelerated titration designs were evaluated by computer simulation both with and without intra-patient dose escalation.

### ***Model-based analysis***

Since accelerated titration designs utilize graded toxicity results and multi-course treatment results, the information yield can be greater than for conventional or CRM designs. The model used by Simon et al. <sup>35</sup> was based on measuring the worst toxicity experience by each patient during each course of treatment. That is, the model does not consider separate toxicity for each organ system separately, but takes the maximum over the organ systems and records that worst toxicity separately for each course of treatment for each patient. The toxicity for patient  $i$  in course  $j$  is determined by

$$\log (d_{ij} + \alpha D_{ij}) + \beta_i + \varepsilon_{ij} \quad (2)$$

where  $d_{ij}$  denotes the dose received by patient  $i$  in course  $j$ , and  $D_{ij}$  denotes the cumulative dose received by patient  $i$  up to but not including course  $j$ . For the first course,  $D_{ij}$  is zero for all patients.  $\alpha$  is a cumulative toxicity parameter and  $\alpha = 0$  represents no cumulative toxicity. All logarithms are natural logarithms. The  $\beta_i$  terms represent inter-patient variability in toxic effects. The  $\beta_i$  term is the same for all courses of treatment of patient  $i$  but its value differs among patients. The  $\beta_i$  values are taken as independent draws from a normal distribution with zero mean and variance  $\sigma_\beta^2$ . Hence, the model has a single parameter ( $\sigma_\beta^2$ ) that reflects the amount of inter-patient

variability in susceptibility to toxicity. The  $\varepsilon_{ij}$  terms are the random variations that reflect the uncontrolled sources of variation other than dose that influence the toxic response for a given patient. These are taken as independent draws from a normal distribution with zero mean and variance  $\sigma_{\varepsilon}^2$ .

In addition to the three parameters  $\alpha$ ,  $\sigma_{\beta}^2$  and  $\sigma_{\varepsilon}^2$ , there are also several parameters for converting the quantitative value of (2) into a graded level of toxicity. Values of expression (2) less than  $K_1$  correspond to less than intermediate toxicity. Values between  $K_1$  and  $K_2$  correspond to intermediate toxicity, values between  $K_2$  and  $K_3$  correspond to dose-limiting toxicity, and values greater than  $K_3$  correspond to life-threatening toxicity. If one doesn't wish to distinguish DLT from life-threatening toxicity, then only  $K_1$  and  $K_2$  are needed. So there are 5 - 6 parameters to be estimated from the data. This model is a generalization of the  $K_{max}$  model of Sheiner et al.<sup>33</sup>, and of the model of Chou and Talalay<sup>7</sup>.

Given the data of the grade of toxicity (worst over organ systems) for each course of each patient, the method of maximum likelihood is used to estimate the model parameters. Splus software for fitting the parameters is available at <http://linus.nci.nih.gov/~brb>. That web site also contains an Excel macro for managing dose assignments to patients during Accelerated Titration Design trials. The macro assists investigators in quality controlling the dose assignment process and provides a convenient way of recording dose assignments in a systematic manner that makes the data available for subsequent analysis.

Simon et al.<sup>35</sup> fit the model (2) to data from 20 phase I trials. Only 3 of the trials showed any evidence of cumulative toxicity ( $\alpha > 0$ ). The estimates of  $\alpha$  for the other trials were zero or very close to zero. The trials varied substantially in the other parameters and thus provide a broad range of experience for evaluation of the accelerated titration designs.

### ***Evaluation of performance***

Simon et al.<sup>35</sup> evaluated the performance of accelerated titration designs by simulating phase I data based on the twenty sets of parameters estimated from the twenty real trials that they studied. For each of the twenty sets

of parameters, they generated data for 1000 phase I trials and applied each of their designs to the simulated data.

Figure 1 shows the average number of patients per trial utilized by each of the designs. For each design, the average is taken over the same 20,000 simulated data sets generated from the sets of parameters derived from the 20 actual trials analyzed. Results for eight designs are shown. Designs 1-4 are as described above. The designs labeled with B utilize intra-patient dose escalation if the toxicity in the previous course is less than intermediate. Designs labeled with A do not permit intra-patient dose escalation.

Design 1A corresponds to the standard design, although it does not use Fibonacci dose steps. Design 1B is the standard design augmented to permit intra-patient dose escalation. As can be seen in Figure 1, the average number of patients is much greater for the standard design 1A or 1B than for any of the accelerated titration designs. The average number of patients is somewhat less for designs 3 and 4 that use double dose steps compared to design 2. Although the average differences are not great, the differences for individual trials can be. That is, for a trial in which the starting dose is very low relative to the dose at which intermediate toxicity is expected, designs 2 and 3 will require substantially fewer patients.

Figure 1 also shows the average number of patient cohorts utilized by each design. The average is lowest for designs 3 and 4 that use double dose steps. Although the difference in average number of cohorts is not large, the difference in average time to complete the trials will be much shorter for designs 2 - 4 if patients are not instantaneously available since the accelerated phase of those designs requires only one patient per cohort.

Figure 2 shows the average number of patients experiencing each level of toxicity as their worst toxicity during their treatment on the trial. With the standard design, an average of 23 patients experience less than intermediate toxicity (labeled “no toxicity” in the figure). These patients are under-treated. For design 2B the average number of under-treated patients is about 8 and for designs 3B and 4B the number is less than 5. This major reduction in the number of under-treated patients is achieved with very small increases in the average number of patients experiencing DLT or unacceptable toxicity with the accelerated titration designs.

The accelerated titration designs without intra-patient dose escalation, 2A, 3A and 4A, performed quite well with regard to reduction in average number of patients and reduction of number of under-treated patients. They do not provide patients accrued early in the trial a full opportunity to be treated at a therapeutic dose, however. They are also less effective in situations where inter-patient variability in susceptibility to toxicity is large. These designs

may be attractive, however, when there is concern about cumulative toxicity. It is worth noting, in this regard, that analysis of the 20 phase I trials used for evaluation of these designs revealed no evidence of ill effect from intra-patient dose escalation and lead the investigators to conclude that “cumulative toxicity does not appear to be a valid reason to prohibit intra-patient dose escalation, as it occurs rarely”<sup>2</sup>.

Accelerated titration designs can dramatically reduce the number of patients accrued to a phase I trial. They can also substantially shorten the duration of the phase I trial. They provide much greater information than other designs with regard to cumulative toxicity, inter-patient variability and steepness of the dose-toxicity curve. They also provide all patients entered in the trial a maximum opportunity to be treated at a therapeutic dose.

## **Pharmacokinetically Guided Dose Escalation**

An entirely different approach to the problem of safely accelerating the dose in phase I studies was proposed by Collins et al.<sup>10</sup>. They presented a retrospective analysis of anti-cancer agents which demonstrated that, for the most part, toxicity was not a function of administered drug dosage, but, rather, was a function of AUC, the area (C x T) under the curve of plasma drug concentration (C) measured over time of exposure (T). Therefore, they proposed a “pharmacokinetically guided dose escalation” (PGDE) scheme that involved targeting the AUC associated with the mouse LD<sub>10</sub> (rather than the MELD<sub>10</sub> itself).

### ***Initial PGDE proposal of Collins et al.<sup>10</sup>***

The initial PGDE scheme of Collins et al.<sup>10</sup> involved escalating to an MTD by targeting a maximal tolerated AUC, and proceeded as follows:

- 1) Determine the mouse LD<sub>10</sub>, and the associated mouse AUC, of the new agent
- 2) Treat the initial cohort of three patients at 1/10 MELD<sub>10</sub>, as is standard, and measure the average (human) AUC over this cohort of patients
- 3) Escalate the doses for subsequent cohorts of three patients according to the distance to the target AUC (that which is associated with the MELD<sub>10</sub>), according to one of the following two rules:
  - i) First escalation step increases the initial dose by a factor equal to the square root of the ratio of the

target AUC to the AUC associated with the initial dose, and subsequent escalation steps follow the Fibonacci scheme

- ii) Escalation steps are by a factor of two until the AUC is 40% of the target AUC, and subsequent escalation steps follow the Fibonacci scheme

Retrospective analyses by Collins et al. <sup>10</sup> indicated that the sample sizes of phase I trials could be reduced by 20% - 50% by utilizing this pharmacokinetically guided dose escalation (PGDE) scheme.

The efficiency of PGDE relies on the assumption that drug toxicity is really a function of drug AUC, and that equivalent AUC for human and mouse will result in equivalent toxicity. Furthermore, the underlying assumption is that the mouse LD<sub>10</sub> roughly equals the human MTD, both doses measured as a function of body surface area (in mg/m<sup>2</sup>) because, in general, the two doses yield roughly equivalent AUC levels for the two species. However, as noted by Collins et al. <sup>10</sup>, there are important exceptions to this rule. For example, the MTD of doxorubicin in humans is five-fold higher than the MELD<sub>10</sub> because the clearance rate of doxorubicin is much higher in man than in the mouse, leading to a much smaller AUC in man for the equivalent dose. This sort of situation leads to a striking advantage for PGDE, since the smaller than expected AUC for the first dose will result in escalation of the initial dose step(s). Other situations, also noted by Collins et al. <sup>10</sup>, on the other hand, may lead to problems for PGDE. For some drugs, there is a drug concentration threshold for action and a necessary minimum exposure time above that threshold. For such drugs, the relation between mouse toxicity and human toxicity is complicated by the fact that, in general, the smaller species experiences a higher initial drug concentration and a shorter half-life <sup>15</sup>. Thus, if the threshold is high and the necessary exposure time short, the mouse may experience much more serious toxicity for equivalent AUC (or equivalent dose). Likewise, if the threshold is low and the necessary exposure time long, the human may experience much more serious toxicity. For other drugs, there is a marked difference between the two species in target cell sensitivity, again rendering mouse drug dose or AUC non-predictive of the human MTD.

### ***Reception and status of PGDE***

The European Organization for Research and Treatment of Cancer (EORTC) published a generally positive review of PGDE <sup>13</sup>. However, the EORTC <sup>13</sup> also re-iterated the cautions given by Collins et al. <sup>10</sup>, and it added

some of its own, in particular, the following:

- 1) For pharmacokinetic measures to translate across species, the pre-clinical conditions (in particular, the route of administration) should match the anticipated clinical conditions
- 2) Metabolism of active (and toxic) agents in humans, but not in mice, may complicate the use of the AUC (for the drug alone) to predict toxicity across the species
- 3) The plasma drug concentration is really a surrogate measure for target tissue drug concentration, so if the relationship between the two is not equivalent for the two species, it may not be predictive of toxicity
- 4) Anti-metabolite toxicity is not well predicted by AUC (or by dose)

Despite the above cautions, the EORTC review<sup>13</sup> of PGDE was positive.

The EORTC review<sup>13</sup> concluded by noting that “the worldwide experience to date reveals no example in which the use of PGDE would introduce a greater risk to patients than the procedures currently employed. In many cases the new procedures would lead to a more efficient dose escalation.” The EORTC re-iterated the initial Collins PGDE scheme and proposed that it be evaluated prospectively.

A later review by Collins et al.<sup>9</sup> reviewed the role of prospective application of PGDE concepts to eight consecutive NCI-sponsored trials. For three of the drugs, development was reported to be too far advanced for impact from the PGDE project. For two (merbarone and deoxyspergualin), pre-clinical data demonstrated that a continuous infusion schedule would reduce toxicity and allow use of a substantially higher initial phase I dose (for HMBA pre-clinical data also enabled a higher initial dose), which Collins et al.<sup>9</sup> claimed to be an extension of the PGDE concepts. For the remaining two (flavone acetic acid and pirozantrone), phase I dose escalation was accelerated according to the PGDE scheme. Collins et al.<sup>9</sup> also noted the use of a related concept, using AUC measurements to individually adapt patient dosage, in particular, in cases where there is wide variation in AUC among patients receiving the same dose. They noted the use of this “adaptive control” approach in the development of regimens for etoposide and HMBA.

As PGDE continued to be used, problems arose, some of which were successfully addressed, and some not. Gianni et al.<sup>16</sup> reported on the discovery, in the course of a phase I trial of I-Dox, that it metabolized, in the human but not in the mouse, to the active and toxic agent I-Doxol, which attained plasma concentrations 10-fold those of the original drug. PGDE was eventually utilized successfully in this trial, based on the combined AUC of I-Dox

plus I-Doxol, in the human, equated to the AUC of I-Dox in the mouse. Fuse et al.<sup>15</sup> argued that new agents can often be classified into one of two types. For type I drugs (including alkylating agents and some anti-tumor antibiotics), toxicity is a function of AUC, and therefore, PGDE can be advantageously used. For type II drugs (anti-metabolites and vinca alkaloids), toxicity is a function of exposure time, rather than AUC, and the use of PGDE is not possible.

Another situation which makes it impossible to use PGDE is the presence of large inter-patient variability in AUC, for the same administered dose. The advantage of using AUC, as opposed to dose, is based on the assumption that inter-species variation in AUC is high (in particular, between the mouse and the human), but intra-species variation (for both mice and men) is low, so that equivalent doses will result in predictable AUC levels for the two species (although not equivalent). Conley et al.<sup>11</sup> reported that for HMBA, it was found that the variability of AUC was very high among patients on the phase I trial receiving equivalent doses. In this case, the use of adaptive control was possible, since AUC, although variable, proved to be predictive of toxicity. On the other hand, for CI-941, Foster et al.<sup>14</sup> reported that not only was the AUC quite variable, but also AUC was no more predictive of toxicity than was administered dose. For this drug, neither PGDE nor adaptive control, based on individual AUC measures, would be useful.

Collins<sup>8</sup> sums up the present state of PGDE. Despite encouraging reports on its success in the United States, Europe, and Japan, investigators, in the end, find the requirement of real-time pharmacokinetic monitoring to be a drawback. Collins<sup>8</sup> concludes that although such pharmacokinetics could prove useful, because of this attitude “PGDE has failed to be widely accepted, and has generally faded from regular use.” Similarly, Newell<sup>27</sup> noted that the common failure to collect pharmacodynamic data and data relating to biologic efficacy against the intended targets of new agents is seen as a serious failure by the pre-clinical investigators involved in drug development.

## **Phase I designs for non-toxic therapeutics**

Certain types of therapeutics are not expected to be toxic in the dose range used. Some molecularly targeted drugs and therapeutic vaccines are of this type. Conventional phase I designs are not suitable for such drugs because there is no interest in the maximum tolerated dose. Nevertheless, there may be uncertainty about the appropriate dose to use for clinical development. Resolving this uncertainty may not be possible, however, in the context of

small 3 - 6 patient per cohort studies used for cytotoxics<sup>21, 36</sup>. Much larger studies may be required, depending on the specific objectives.

### ***Pharmacokinetics designs***

A pharmacokinetics based design generally can be accomplished with a limited number of subjects. One determines a target serum concentration based on pre-clinical or ex-vivo studies. For molecularly targeted drugs, the target concentration is chosen to maximally inhibit the target. The phase I trial then includes  $n$  patients for each of several dose levels. The serum concentration of the active metabolite is measured and the dose chosen that best achieves the target concentration. The target concentration is often a steady state level or a concentration integrated over time (CxT). In some cases the target concentration may be determined based on ex-vivo studies using tissue obtained from subjects on the trial but prior to treatment.

The sample size  $n$  may be derived in the following manner. Suppose one wishes to estimate the mean concentration associated with each dose so that the estimate is within  $100\gamma\%$  of the true mean with high confidence  $1 - \alpha$ . If the serum concentration measurements are normally distributed with constant coefficient of variation ( $cv$ ) and if the coefficient of variation is known, then the required sample size is:

$$n = (z_{1-\alpha/2} cv / \gamma)^2 \quad (3)$$

where  $z_{1-\alpha/2}$  is the  $100(1 - \alpha/2)$ 'nd percentile of the standard normal distribution. A confidence level of 90% corresponds to the  $z$  value of 1.645. If the coefficient of variation is 0.5 and we want to be within 25% of the mean ( $\gamma = 0.25$ ), then we obtain  $n = 11$  patients per dose level. If the accuracy of the assay is greater and there is little inter-patient variability in pharmacokinetics, then we might have a smaller coefficient of variability. If  $cv = 0.25$ , then with the other parameters the same as above, the formula indicates that only 3 patients per dose level are required. Formula (3) is actually an under-estimate of the number of patients required in most circumstances because it assumes that the coefficient of variation is known and does not account for the variability in estimation of the  $cv$  from the data. In general the  $z_{1-\alpha/2}$  term should be replaced by the corresponding percentile of the  $t$  distribution with



degrees of freedom equal to the  $(n - 1)$  times the number of dose levels to be studied.

For some clinical trials, even a reasonable estimate of the coefficient of variation may not be known in advance. In such a case the  $CV$  should probably be estimated based on an initial cohort of patients treated at a fixed dose. After estimating the  $CV$ , the sample size per cohort can be determined from equation (3) for use with the subsequent patients. Since toxicity is not expected, it is best to randomize subsequent patients among the dose levels to be studied.

### ***Minimal biologically active dose***

One may define biological activity based on inhibition of a molecular target, or based on an immunogenic response, and attempt to identify the smallest dose that is biologically active. Table 3 shows the probability of no biological response in  $n$  patients as a function of the true response probability. If one wants a dose at which the response probability is at least 30%, then after observing no responses in 7 patients it would be appropriate to escalate to the next dose level. Simon's optimal two-stage phase II designs can also be used to distinguish a response probability of some uninteresting level  $p_0$  (e.g. 0.05) from a promising level  $p_1$  (e.g. 0.30)<sup>34</sup>. Unless  $p_1$  is much greater than  $p_0$ , however, the required number of patients will be much larger than for cytotoxic phase I trials (see Table 4).

Korn et al.<sup>21</sup> defined a sequential procedure for finding a biologically active dose, although not necessarily the minimal active dose. During an initial accelerated phase they treat one patient per dose level until a biological response is seen. After the first response is seen, they treat cohorts of 3 - 6 patients per dose level. With 0 - 1 biological responses among the 3 patients at a dose level, they escalate to the next level for the next cohort of patients. With 2 or 3 responses out of the 3 patients, they expand the cohort to a total of 6 patients. With 5 or 6 biological responses out of the 6 patients, they declare that dose level to be the biologically active level and terminate the trial. With fewer than 5 biological responses out of the 6 patients, a new cohort of 3 patients is accrued at the next higher dose level, etc. Korn et al.<sup>21</sup> describe some of the statistical properties of this sequential design.

### ***Determining the presence of a dose-response relationship and characterizing the relationship***

Trying to determine whether there is a relationship between dose and biological response involves comparing response rates or response distributions for patients at different dose levels. Such trials, if designed properly, require larger sample sizes. For example, suppose that biological response is binary and one wishes to plan a study of two dose levels and test whether the biological response rates differ at the two levels. If the true response probabilities at the two dose levels are 50% and 90%, then 20 patients treated at each dose level are required for a one-sided statistical significance of 0.10 and a statistical power of 0.90<sup>21</sup>. Larger sample sizes are required to detect smaller differences. Using more than two dose levels allows one to treat somewhat fewer patients at each dose level, but the total number of patients required to detect a dose-response relationship will actually be much larger than if only two dose levels are tested. This is because the two most extreme dose groups are the most informative for detecting a dose-response relationship.

Trying to characterize the shape of the dose-biological response relationship or finding an optimum biologic dose (OBD) is even a more ambitious objective than a two dose comparison of biological response rates. It is rarely practical in a phase I study unless there is an accurate quantitative assay of biological response with little intra-patient or inter-patient variability in assay results.

Trials utilizing biological response endpoints are also complicated by issues of assay adequacy and access to biological tissues. Because of the difficulties of accessing tumor tissue, some studies have used normal tissue in which the molecular target is highly expressed<sup>1</sup>. Thus, one strategy for phase I study of molecularly targeted drugs is to compare dose levels with regard to biological response in accessible normal tissue using an optimized highly reproducible assay. The use of normal tissue may serve to reduce inter-patient variability.

If use of normal tissue for assessing biological response is not acceptable or if a highly reproducible assay is not available, trying to characterize an OBD is probably not feasible. In such cases it would probably be better to optimize the dose level utilizing clinical response as the endpoint. It would probably take more patients and more time to characterize the OBD than to compare dose levels with regard to clinical response. Such studies can either use tumor shrinkage or time to progressive disease as the clinical endpoint. The studies are best conducted as randomized trials but the type I error level  $\alpha$  need not be set stringently at the conventional 5% level. Several authors such as Budde and Bauer<sup>4</sup> and Chen and Simon<sup>6</sup> have developed designs that can be used for clinical trials

with dose-response objectives.

## Further considerations

Several other alternative phase I designs have been proposed for special situations. For drugs with variable dose effect based on a patient baseline characteristic (initial white blood count, in particular), Mick and Ratain<sup>26</sup> suggested using a dose-toxicity model, incorporating this additional variable, to define both the MTD and the dose-escalation schema. For phase I studies of drug combinations, Korn and Simon<sup>22</sup> point out that there may be a wide variety of combined MTD's, involving different drug proportions. They provide guidance in arriving at a favorable combination, from a dose-intensity perspective, as well in designing the combined dose-escalation schema.

Phase I studies in children are generally performed after an adult MTD has been established, and dose-escalation begins at 80% of the adult dose, to minimize under-treatment<sup>37</sup>. In the past, phase I studies have routinely excluded elderly patients due to the assumption that they are inherently more susceptible to toxicity. Cascinu et al.<sup>5</sup> review the results of a comparison of toxicity seen in 120 elderly patients, compared with that seen in 120 non-elderly patients with similar clinical features and receiving the same chemotherapeutic regimens. They report that the chemotherapeutic regimens, for both groups, yield similar benefits and similar toxicities. They conclude that chronologic age is a weak predictor of either toxicity or failure to respond.

Except for phase I studies of non-toxic therapeutics, which we have discussed at length, there is usually little attempt to assess the efficacy of the therapy in the phase I trial. Sometimes the MTD cohort is expanded to approximately 10 patients, to further assess toxicity. In these cases, care should be taken to not over-interpret any responses seen, or, more importantly, lack of response. Phase II trials of efficacy are significantly larger than 10 patients, and still give only crude indications of response rate. Moreover, 10 patients without response is generally insufficient evidence upon which to reject the potential efficacy of an agent. Most importantly, the patient population of phase I trials is generally not as favorable as that of phase II trials with respect to the likelihood of seeing tumor response.

The last decade has witnessed a dramatic change in the design and practice of phase I trials. A study of the phase I trials conducted at M.D. Anderson Cancer Center between 1991 and 1993<sup>38</sup> concluded that all used standard designs, with 23% of the patients treated at less than 50% of the determined MTD. In contrast, all participants in a

1998 colloquium on new phase I designs<sup>12</sup> agreed that they no longer used standard phase I designs routinely, that dose escalation with one-patient cohorts in the initial stages of phase I trials had become frequent and was apparently safe, and that all of the commonly used newer methods are generally preferable to the standard phase I design.

## Bibliography

1. Albanell J, Rojo F, Averbuch S, et al.: Pharmacodynamic studies of the epidermal growth factor receptor inhibitor ZD1839 in skin from cancer patients: histopathologic and molecular consequences of receptor inhibition. Journal of Clinical Oncology 20:110, 2002
2. Arbusk SG: Workshop on phase I study design—Ninth NCI/EORTC New Drug Development Symposium. Annals of Oncology 7:567, 1996
3. Babb J, Rogatko A and Zacks S: Cancer phase I clinical trials: Efficient Dose Escalation with Overdose Control. Statistics in Medicine 17:1103, 1998
4. Budde M and Bauer P: Multiple test procedures in clinical dose finding studies. Journal of the American Statistical Association 84:792, 1989
5. Cascinu S, Del Ferro E and Catalano G: Toxicity and therapeutic response to chemotherapy in patients aged 70 years or older with advanced cancer. American Journal of Clinical Oncology 19:371, 1996
6. Chen TT and Simon R: A multiple decision procedure in clinical trials. Statistics in Medicine 13:431, 1994
7. Chou TC and Talalay P: Generalized equations for the analysis of inhibitions of Michaelis-Menten and higher-order kinetic systems with two or more mutually exclusive and nonexclusive inhibitors. European Journal of Biochemistry 115:207, 1981
8. Collins JM: Innovations in phase I trial design: Where do we go next? Clinical Cancer Research 6: 3801, 2000
9. Collins JM, Grieshaber CK and Chabner BA: Pharmacologically guided phase I clinical trials based upon preclinical drug development. Journal of the National Cancer Institute 82:132, 1990
10. Collins JM, Zaharko DS, Dedrick RL, et al.: Potential roles for preclinical pharmacology in phase I clinical trials. Cancer Treatment Reports 70:73, 1986
11. Conley BA, Forrest A, Egorin MJ, et al.: Phase I trial using Adaptive Control dosing of hexamethylene bisacetamide. Cancer Research 49:3436, 1989
12. Eisenhauer EA, O'Dwyer PJ, Christian M, et al.: Phase I clinical trial design in cancer drug development. Journal of Clinical Oncology 18:684, 2000
13. EORTC Pharmacokinetics and Metabolism Group: Pharmacokinetically guided dose escalation in phase I clinical trials: Commentary and proposed guidelines. European Journal of Clinical Oncology 23:1083, 1987
14. Foster BJ, Newell DR, Graham MA, et al.: Phase I trial of the anthrapyrazole CI-941: Prospective

- evaluation of a pharmacokinetically guided dose-escalation. European Journal of Cancer 28:463, 1992
15. Fuse E, Kobayashi T, Inaba M, et al.: Prediction of the maximal tolerated dose (MTD) and therapeutic effect of anticancer drugs in humans: Integration of pharmacokinetics with pharmacodynamics and toxicodynamics. Cancer Treatment Reviews 21:133, 1995
  16. Gianni L, Vigano L, Surbone A, et al.: Pharmacology and clinical toxicity of 4'-Iodo-4'-deoxydoxorubicin: An example of successful application of pharmacokinetics to dose escalation in phase I trials. Journal of the National Cancer Institute 82:469, 1990
  17. Goodman SN, Zahurak ML and Piantadosi S: Some practical improvements in the continual reassessment method for phase I studies. Statistics in Medicine 14:1149, 1995
  18. Grieshaber CK and Marsoni S: The relation of preclinical toxicology to findings in early clinical trials. Cancer Treatment Reports 70:65, 1986
  19. Korn EL, Arbuck SG, Pluda JM, et al.: Clinical trial designs for cytostatic agents: Are new approaches needed? Journal of Clinical Oncology 19:265, 2001
  20. Korn EL, Midthune D, Chen TT, et al.: A comparison of two phase I trial designs, Statistics in Medicine 13:1799, 1994
  21. Korn EL, Rubinstein LV, Hunsberger SA, et al.: Clinical trial designs for cytostatic agents and agents directed at novel molecular targets. In Buolamwini JK and Adjei AA (eds.) Novel Anticancer Drug Protocols (Methods in Molecular Medicine Series), Humana Press, in press 2003
  22. Korn EL and Simon R: Using the tolerable-dose diagram in the design of phase I combination chemotherapy trials. Journal of Clinical Oncology 8:374, 1990
  23. Leventhal BG and Wittes RE: Research Methods in Clinical Oncology, New York, Raven Press, 1988
  24. Lindley DV: Bayesian inference In Kotz S and Johnson NL (eds): Encyclopedia of Statistical Sciences, Volume 1, New York, John Wiley & Sons, p 197, 1982
  25. Mahmood I and Balian JD: The pharmacokinetic principles behind scaling from preclinical results to phase I protocols. Clinical Pharmacokinetics 36:1, 1999
  26. Mick R and Ratain MJ: Model-guided determination of maximum tolerated dose in phase I clinical trials: evidence for increased precision. Journal of the National Cancer Institute 85:217, 1993
  27. Newell DR: Pharmacologically based phase I trials in cancer chemotherapy. Hematology/Oncology Clinics of North America 8:257, 1994
  28. O'Quigley J: Estimating the probability of toxicity at the recommended dose following a phase I clinical trial in cancer. Biometrics 48:853, 1992
  29. O'Quigley J and Chevret S: Methods for dose finding studies in cancer clinical trials: A review and results of a Monte Carlo study. Statistics in Medicine 10:1647, 1991
  30. O'Quigley J, Pepe M and Fisher L: Continual reassessment method: a practical design for phase I clinical trials in cancer. Biometrics 46:33, 1990
  31. Potter DM: Adaptive dose finding for phase I clinical trials of drugs used for chemotherapy of cancer.

Statistics in Medicine 21:1805, 2002

32. Schneiderman MA: Mouse to man: statistical problems in bringing a drug to clinical trial. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, p 855, 1967
33. Sheiner LB, Beal SL, and Sambol NC: Study designs for dose ranging. Clinical Pharmacology and Therapeutics 46:63, 1989
34. Simon R: Optimal two-stage designs for phase II clinical trials. Controlled Clinical Trials 10:1,1989
35. Simon RM, Freidlin B, Rubinstein LV, et al.: Accelerated titration designs for phase I clinical trials in oncology. Journal of the National Cancer Institute 89:1138, 1997
36. Simon RM, Steinberg SM, Hamilton M, et al.: Clinical trial designs for the early clinical development of therapeutic vaccines. Journal of Clinical Oncology 19:1848, 2001
37. Smith M, Bernstein M, Bleyer WA, et al.: Conduct of phase I trials in children with cancer. Journal of Clinical Oncology 16:966, 1998
38. Smith TL, Lee JJ, Kantarjian HM, et al.: Design and results of phase I cancer clinical trials: Three-year experience at M.D. Anderson Cancer Center. Journal of Clinical Oncology 14:287, 1996
39. Storer BE: Design and analysis of phase I clinical trials. Biometrics 45:925, 1989
40. Therasse P, Arbuck SG, Eisenhauer EA, et al.: New guidelines to evaluate the response to treatment in solid tumors (RECIST guidelines), Journal of the National Cancer Institute, 92:205, 2000

**Table 1: Dose Escalation Rules for the Standard Phase I Trial**

<b><i>Outcome:</i></b> # DLT / # Pts	<b><i>Action:</i></b> Escalate, suspend, or halt dose escalation
0 DLT out of 3 patients	Escalate dose for next cohort of 3 patients
1 DLT out of 3 patients	Treat next cohort of 3 patients at the same dose
$\geq 2$ DLT out of 3 patients	Halt dose escalation: treat total of 6 patients at previous dose to determine MTD*
1 DLT out of 6 patients	Escalate dose for next cohort of 3 patients
$\geq 2$ DLT out of 6 patients	Halt dose escalation: treat total of 6 patients at previous dose to determine MTD*

\*MTD--the highest dose for which no more than 1 of the 6 treated patients exhibits DLT

**Table 2: Probabilities of Halting or Continuing Dose Escalation for Various Probabilities of DLT Associated with the Dose Level, for the Standard Phase I Design**

True probability of DLT for dose level	.05	.1	.2	.3	.4	.5	.6	.7
Probability of halting dose escalation after accruing either 3 or 6 patients ( $\geq 2$ DLT) <sup>(1)</sup>	.03	.09	.29	.51	.69	.83	.92	.97
Probability of continuing escalation after only 3 patients (0 DLT) <sup>(2)</sup>	.86	.73	.51	.34	.22	.13	.06	.03
Probability of halting escalation after only 3 patients ( $\geq 2$ DLT) <sup>(2)</sup>	.01	.03	.10	.22	.35	.50	.65	.78

- (1) This row gives probabilities of halting dose escalation, at a given dose, if the true probability of DLT for that dose level is as indicated.
- (2) These rows gives probabilities of continuing or halting dose escalation after accruing only 3 patients, at a given dose, if the true probability of DLT for that dose level is as indicated. We see that, in all cases, the cohort will be limited to 3 patients with at least 50% probability, and for the more extreme DLT probabilities (.05 or .7), the cohort will be expanded to 6 patients with less than 20% probability.



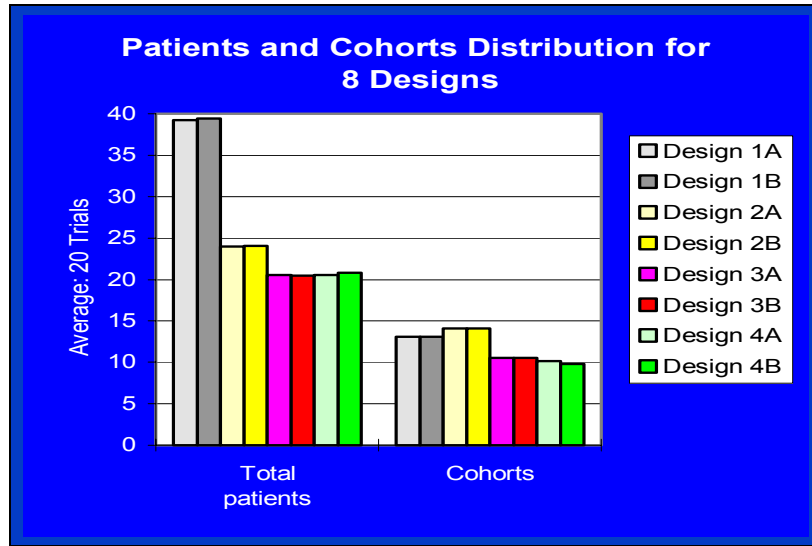
**Table 3: Finding the Minimum Active Dose**

<b>Probability of Biological Response</b>	<b>Number of Patients Treated at Dose</b>	<b>Probability of No Biological Responses</b>
0.20	11	0.09
0.25	9	0.08
0.30	7	0.08
0.40	5	0.08
0.50	4	0.06

**Table 4: Optimal Two-Stage Designs**

<b>Target Response Rate (<math>p_1</math>)</b>	<b>First Stage Sample Size (<math>N_1</math>)</b>	<b>Maximum Sample Size (<math>N</math>)</b>	<b>Number of Responses Required For Activity (A)</b>	<b>Probability of Early Termination</b>
20%	12	37	4	.54
25%	9	24	3	.63
30%	7	21	3	.70
35%	6	12	2	.74

**Figure 1**



**Figure 2**

