Clinical Cancer Research

Seamless Phase I-II Trial Design for Assessing Toxicity and Efficacy for Targeted Agents

Antje Hoering, Mike LeBlanc and John Crowley

Updated Version	Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-10-1262

Cited Articles	This article cites 6 articles, 2 of which you can access for free at: http://clincancerres.aacrjournals.org/content/17/4/640.full.html#ref-list-1

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

Clinical Cancer Research

Seamless Phase I-II Trial Design for Assessing Toxicity and Efficacy for Targeted Agents

Antje Hoering[1], Mike LeBlanc[2], and John Crowley[1]

Abstract

Purpose: The premise for phase I trials for cytostatic agents is different from that of cytotoxic agents. For cytostatic agents, toxicity and efficacy do not necessarily increase monotonically with increasing dose levels, but likely plateau after they reach maximal toxicity or efficacy. Here, we propose a phase I-II trial design to assess both toxicity and efficacy to find the best dose as well as a good dose.

Experimental Design: We propose a 2-step dose-finding trial for assessing both toxicity and efficacy for a targeted agent. The 1st step uses a traditional phase I trial design. This step only assesses toxicity and finds the maximal tolerated dose (MTD). For the 2nd step, we propose a modified phase II selection design for 2 or 3 dose levels at and below the MTD to determine efficacy and evaluate each dose level by both efficacy and toxicity.

Results and Conclusion: Simulation studies are done on several combinations of toxicity and efficacy scenarios to assess the operating statistics of our proposed trial design. We then compare our results with a traditional phase I trial followed by a single-arm phase II trial using the same total sample size. The proposed design does better in most cases than a traditional design using the same overall sample size. This design allows assessing a few dose levels more closely for both efficacy and toxicity and provides greater certainty of having correctly determined the best dose level before launching into a large efficacy trial. *Clin Cancer Res; 17(4); 640–6. ©2010 AACR.*

Background

The main goal in phase I trials for traditional cytotoxic agents is to determine the maximal tolerated dose (MTD). The underlying premise is that both efficacy and toxicity increase monotonically with increasing dose levels. Only toxicity, not efficacy, is monitored during a traditional phase I trial. The standard 3+3 design accrues 3 to 6 patients at a time to a given dose level and then increases the dose level until dose limiting toxicity (DLT) is observed. If 2 or more DLTs are observed in a group of 6 patients at that dose level, dose escalation ceases and the MTD has been exceeded. The highest dose in which no more than 1 DLT in 6 subjects is observed is the MTD. Storer (1) recently reviewed the performance of this and other traditional phase I trial designs.

The premise for phase I trials for cytostatic or targeted agents is generally different. Because the agent is designed to specifically interfere with a molecular pathway directly related to specific characteristics of the tumor, it is hypothesized to be less toxic than a traditional cytotoxic agent.

Toxicity does not necessarily increase with increasing dose levels. Efficacy does not necessarily increase monotonically with increasing dose levels either, but may plateau after it reaches maximal efficacy; higher dose levels past this point no longer yield higher efficacy.

Thus, the goal for dose-finding trials for targeted agents should be to determine the dose level that provides highest efficacy while assuring the safety of that dose level. We refer to this dose as the best dose. A variety of continual reassessment models (CRM) have been proposed for this purpose; see, for example, (2, 3). Hunsberger and colleagues (4) recently proposed a dose escalation trial for targeted therapies similar to the traditional 3+3 phase I trial, but with dose escalation solely based on response, assuming that no significant toxicity will occur. These proposed trial designs address the issue of finding such a dose and have good statistical properties. None of these trial designs seems to have found widespread acceptance in the clinical trials community yet. Here, we propose a phase I-II trial design to assess both toxicity and efficacy to find the best dose, as well as a good dose. In this context, the *best* dose is defined as the dose level that maximizes efficacy while assuring safety, and a *good* dose is defined as a dose level in which efficacy is above a predefined boundary while maintaining safety. Targeted agents are often difficult and expensive to manufacture in larger quantities, and a smaller dose provides economic benefit. Thus, under some circumstances, a *good* dose may even be preferable to the *best* dose. Jain and colleagues (5) recently evaluated several phase I trials for

Authors' Affiliations: [1]Cancer Research and Biostatistics; [2]Fred Hutchinson Cancer Research Center, Seattle, Washington

Corresponding Author: Antje Hoering, Cancer Research and Biostatistics (CRAB), 1730 Minor Ave STE 1900, Seattle, WA 98101-1468. Phone: 206-839-1789; Fax: 206-652-4612; E-mail: antjeh@crab.org

©2010 American Association for Cancer Research.

American Association for **Cancer Research**

targeted agents and found evidence that patients on lower dose levels do not necessarily fare worse.

The proposed design can easily be implemented and interpreted. It allows for extended cohorts of patients at dose levels close to the best dose to more precisely determine toxicity and efficacy of the new agent. In addition, different patient populations may be enrolled to the phase I and phase II portion. Traditionally, the patient population for assessing toxicity is broader than the patient population in which efficacy is first tested.

Phase I-II Trial Design for Targeted Agents

Here, we propose a 2-step dose-finding trial for assessing both toxicity and efficacy for a new targeted agent. Both steps will be implemented in the same protocol to insure seamless continuation. For the 1st step, we use a traditional phase I trial design, such as the 3+3, the accelerated titration, or the CRM model. This step only assesses toxicity and finds the MTD. This step insures that the dose levels at and below the MTD are safe in humans. Even if a new agent is not anticipated to have toxicity and has been shown to be safe in animal models, it is important to be certain of that fact before exposing a large number of humans to a new agent (6).

The goal of the 2nd step is to determine the best dose in terms of efficacy and toxicity as a dose level no larger than the MTD. Great care has to be taken in determining the best efficacy endpoint for this part of the trial. Defining an early efficacy endpoint on the basis of tumor biology for these agents is often difficult. In addition, some of these targeted agents are not necessarily expected to yield sufficient tumor shrinkage to achieve a clinical response by standard response criteria [e.g., Response Evaluation Criteria in Solid Tumors (RECIST)]. One possibility is to use progression free survival at a single time point or disease control rate (clinical response of stable or better).

For this 2nd step, we propose a phase II modified selection design (7) for 2 or 3 dose levels at and below the MTD to determine efficacy and evaluate a dose level for both efficacy and toxicity. We assume that a binary endpoint for efficacy such as the ones discussed above has been determined. We propose to accrue approximately 15 to 20 patients per dose level and assess both toxicity and efficacy for those patients. Each dose level is an arm in our phase II trial. We first evaluate each arm independently for both efficacy and toxicity. We do a simple hypothesis test to determine efficacy and assess the power of the test statistic by determining the probability of passing the efficacy boundary independently in each arm. We also determine how many patients experience a DLT and define a toxicity boundary, which is traditionally 33%. If the percentage of patients experiencing a DLT at a specific dose level (arm) is larger than or equal to the toxicity boundary, this dose level is considered to be too toxic and is not pursued any further. On the other hand, if the percentage of patients experiencing a DLT in a specific arm is lower than the toxicity boundary, we consider this arm as having acceptable toxicity. We next determine the probability of picking the arm with the largest efficacy, while assuring acceptable toxicity and a minimal efficacy level as defined above using a slightly modified methodology of selection designs. This expanded cohort of 15 to 20 patients for 2 or 3 dose levels allows us to get a more precise estimate of toxicity and efficacy and, thus, a higher probability of correctly determining the best dose before launching into a larger trial.

Underlying Model Assumptions and Simulation Studies

We assume that toxicity and efficacy are binary measures. In general, toxicity and efficacy are closely linked. Each dose level has a specific average toxicity and efficacy associated with it. We, thus, simulate the toxicity and efficacy data using a correlated bivariate logistic regression model. The correlation can be measured by a correlation coefficient or an odds ratio relating the 2 endpoints. We chose the odds ratio as a means to measure the correlation as it has better numerical properties and an R-package (VGAM) is readily available (8).

Let the marginal probabilities (for toxicity and efficacy) be logistic and depend on the parameter β. For an observation with covariate vector x, the marginal probabilities are then given by:

$$\Pr(Y = 1) = \frac{\exp(x\beta)}{1 + \exp(x\beta)}$$

Let p_{ij} be the joint probability for toxicity $i = (0,1)$ and efficacy $j = (0,1)$. The odds ratio is defined by $\psi = p_{11}p_{00}/p_{10}p_{01}$. For a description of bivariate odds ratio models, see (9). The joint probability p_{11} can be expressed in terms of the marginal probabilities p_1 and p_2 as follows (10):

$$p_{11} = \begin{cases} 1/2(\psi - 1)^{-1}\left(a - \sqrt{a^2 + b}\right) & \text{for} \quad \psi \neq 0 \\ p_1 p_2 & \psi = 0 \end{cases}$$

where $a = 1 + (p_1 + p_2)(\psi - 1)$ and $b = -4\psi(\psi - 1)p_1 p_2$, and p_1 and p_2 denote the marginal probabilities for toxicity and efficacy, respectively.

For our simulation studies, we use 6 dose levels, which is a commonly used number of dose levels for early therapeutic studies. We assume that the dose-response curve is monotonically increasing with increasing dose and remains constant after a critical dose is reached. The window of the 6 dose levels examined may include different parts of that dose-response curve. We distinguish 3 types of efficacy scenarios. As discussed above, efficacy may be measured in different ways depending on the underlying mechanism of the agent of interest. Here, we refer to all the efficacy measures loosely as response measures, keeping in mind, however, that the actual efficacy measure may be different from the traditionally defined response. Figure 1 depicts the 3 response scenarios as a function of
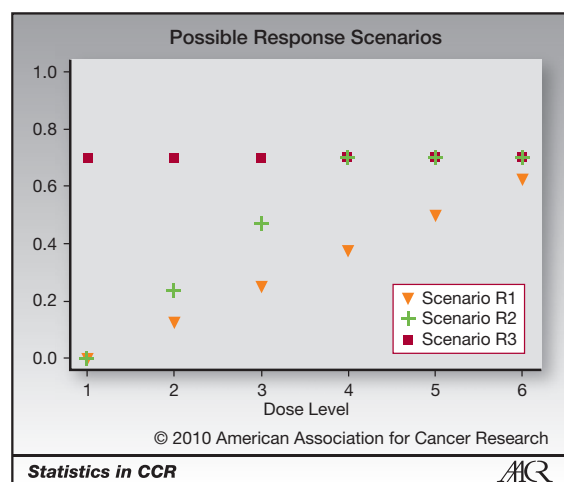
Figure 1. Possible response scenarios as a function of dose level. The marginal probabilities as a function of dose level are plotted.

Figure 2. Possible toxicity scenarios as a function of dose level. The marginal probabilities as a function of dose level are plotted.

dose level. Response scenario R1: This scenario assumes a continuous increase in response with increasing dose level within the dose levels considered. In this case, the leveling-off could occur outside the dose ranges considered. Response scenario R2: In scenario 2, we assume an increase in response for the first 4 dose levels after which it levels off. Response scenario R3: Scenario 3 describes the scenario in which the response is independent of the dose level within the range considered.

Similarly, we assume 3 types of toxicity scenarios; although for the scenario with monotone increase in toxicity, we consider 2 different slopes, so that there is a total of 4 toxicity scenarios. More specifically, these scenarios are: toxicity scenario T1: Scenario 1 assumes that toxicity increases until a maximum toxicity is achieved after which it levels off; toxicity scenarios T2 and T3: Scenarios T2 and T3 assume that toxicity increases monotonically with dose level, in which the increase is steeper for T2 than T3; toxicity scenario T4: Finally, scenario T4 assumes negligible toxicity. The scenarios are illustrated in Fig. 2.

Based on these 4 toxicity scenarios and 3 response scenarios, there are 12 possible combinations of scenarios. The *best* dose is defined as the one that maximizes efficacy while maintaining acceptable toxicity and a minimal efficacy; that is, the rate of dose limiting toxicities is below the toxicity limit, and efficacy passes the efficacy boundary. For each of the response and toxicity scenario combinations, the best dose levels by efficacy and toxicity are summarized in Table 1. In addition, we define a *good* dose level as a dose level with acceptable toxicity and efficacy passing the efficacy boundary.

In our simulation studies, we determined the probability of correctly identifying the MTD in the phase I trial using a traditional 3+3 trial design. A CRM or accelerated titration design could also be used for this step. Table 2 summarizes
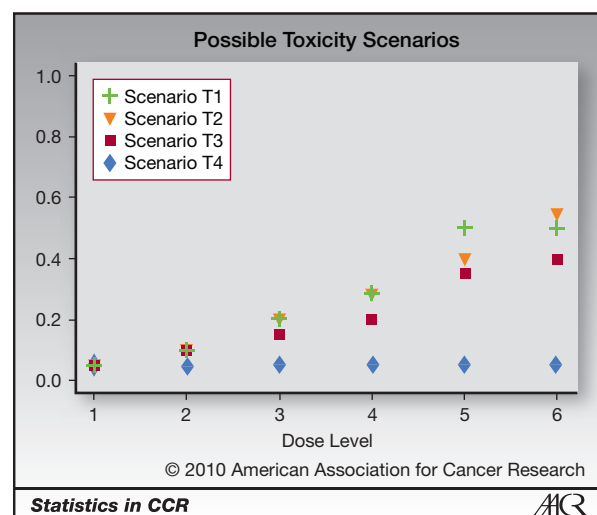
the results of our simulation studies for the phase I portion of the trial. We used 1,000 simulations.

For the phase I trial, 3 different outcomes for each of the toxicity scenarios can be distinguished: the MTD is correctly determined; the MTD is too large; or the MTD is too low. In our simulation studies for the phase II portion, we determine the power of the efficacy test, the probability of the doses being tested to be too toxic, and the probability of correctly determining the best dose. We randomize 40 patients to 2 dose levels, the dose level determined by the phase I part (arm 1) and the dose level immediately below the MTD (arm 2). The hypothesis test for response used in this example tests H0: $P = 0.05$ versus HA: $P = 0.30$. The toxicity limit in our simulations is defined to be 33%. As an example, Table 2 summarizes these results if the MTD was correctly determined in the phase I trial. In our simulation studies, arm 1 is chosen if the toxicity is below the toxicity boundary, if the efficacy is above the efficacy boundary, and if the observed efficacy is larger than the efficacy in arm 2. Arm 2 is chosen if the toxicity is below the toxicity boundary, if the efficacy is above the efficacy boundary, and if the observed efficacy is larger than or equal to the efficacy in arm 1. These 2 probabilities do not add up to 1, as neither arm is chosen if the toxicity is too high or the efficacy is not large enough.

We also compare our results with a combination of the same phase I trial and a traditional single-arm phase II trial at the dose level determined by the phase I trial. We use the same total sample size and determine the probability of correctly picking the best dose level and a good dose level using the definitions above. We evaluate the overall probability of picking a good and best dose level using our proposed design as a sum of the probabilities of the different possible ways to select a *good* or *best* dose level (Fig. 3).

Table 1. Phase I simulation results

Efficacy Scenario	Toxicity Scenario	MTD	Best Dose Level	Good Dose Level	Probability Dose = 0	Probability Dose = 1	Probability Dose = 2	Probability Dose = 3	Probability Dose = 4	Probability Dose = 5	Probability Dose = 6	Average No.
R1	T1	4	4	4	0.02	0.10	0.25	0.29	0.27[a]	0.05	0.01	16
R1	T2	4	4	4	0.03	0.08	0.26	0.30	0.23[a]	0.08	0.02	16
R1	T3	4	4	4	0.03	0.10	0.18	0.21	0.31[a]	0.13	0.05[a]	17
R1	T4	6	6	4–6	0.04	0.02	0.02	0.02	0.03	0.02	0.84[a]	19
R2	T1	4	4	3,4	0.03	0.09	0.25	0.31	0.26[a]	0.05	0.01	16
R2	T2	4	4	3,4	0.03	0.10	0.25	0.28	0.24[a]	0.09	0.01	17
R2	T3	4	4	3,4	0.03	0.09	0.18	0.18	0.31[a]	0.15	0.06	18
R2	T4	6	4–6	3–6	0.03	0.03	0.02	0.04	0.02	0.03	0.84[a]	19
R3	T1	4	1–4	1–4	0.02	0.10	0.26	0.28	0.29[a]	0.04	0.01	16
R3	T2	4	1–4	1–4	0.03	0.09	0.27	0.28	0.22[a]	0.10	0.01	16
R3	T3	4	1–4	1–4	0.03	0.10	0.17	0.22	0.30[a]	0.12	0.07	18
R3	T4	6	1–6	1–6	0.02	0.03	0.02	0.03	0.02	0.02	0.85[a]	19

NOTE. The probability of selecting a specific dose level as a function of efficacy and toxicity. As a reference, the MTD and good and best dose levels for each toxicity and/or efficacy scenario are also given. The best dose is defined as the dose that maximizes response, whereas toxicity is at or below MTD. A good dose is defined as a dose with toxicity at or below the MTD and response $\geq R_A$. The last column gives the average sample size.

[a]The probability of selecting the MTD.

**Table 2.** Phase II simulation results

Efficacy Scenario	Toxicity Scenario	Dose Level Arm 1 (Arm 2)	Power (Arm 1)	Power (Arm 2)	Probability $\geq$ 33% of Patients Having a DLT in Arm 1	Probability $\geq$ 33% of Patients Having a DLT in Arm 2	Probability of Picking the Best Dose Level on the Basis of Efficacy and Toxicity
R1	T1	4 (3)	0.97	0.77	0.30	0.07	0.48
R1	T2	4 (3)	0.98	0.76	0.34	0.08	0.46
R1	T3	4 (3)	0.98	0.77	0.09	0.03	0.66
R1	T4	6 (5)	1.00	1.00	0.00	0.00	0.73
R2	T1	4 (3)	1.00	1.00	0.33	0.08	0.61
R2	T2	4 (3)	1.00	1.00	0.33	0.08	0.60
R2	T3	4 (3)	1.00	1.00	0.08	0.02	0.83
R2	T4	6 (5)	1.00	1.00	0.00	0.00	1.00
R3	T1	4 (3)	1.00	1.00	0.32	0.09	0.92
R3	T2	4 (3)	1.00	1.00	0.34	0.08	0.93
R3	T3	4 (3)	1.00	1.00	0.08	0.03	0.97
R3	T4	6 (5)	1.00	1.00	0.00	0.00	1.00

NOTE. Design properties if MTD is correctly determined. Arm 1 is at the MTD and arm 2 is the dose level right below the best dose.

Results and Discussion

Table 1 summarizes the 12 toxicity and efficacy scenario combinations and their respective MTD, *best*, and *good* dose levels, as defined above. For some scenarios, there is only 1 *best* dose and 1 *good* dose, whereas for others, several or even all dose levels can be considered *good*. In the scenarios R1T1, R1T2, and R1T3, level 4 is the MTD and the only level that crosses the efficacy boundary. On the other extreme is scenario R3T4 in which all levels are considered safe and all levels cross the efficacy boundary.
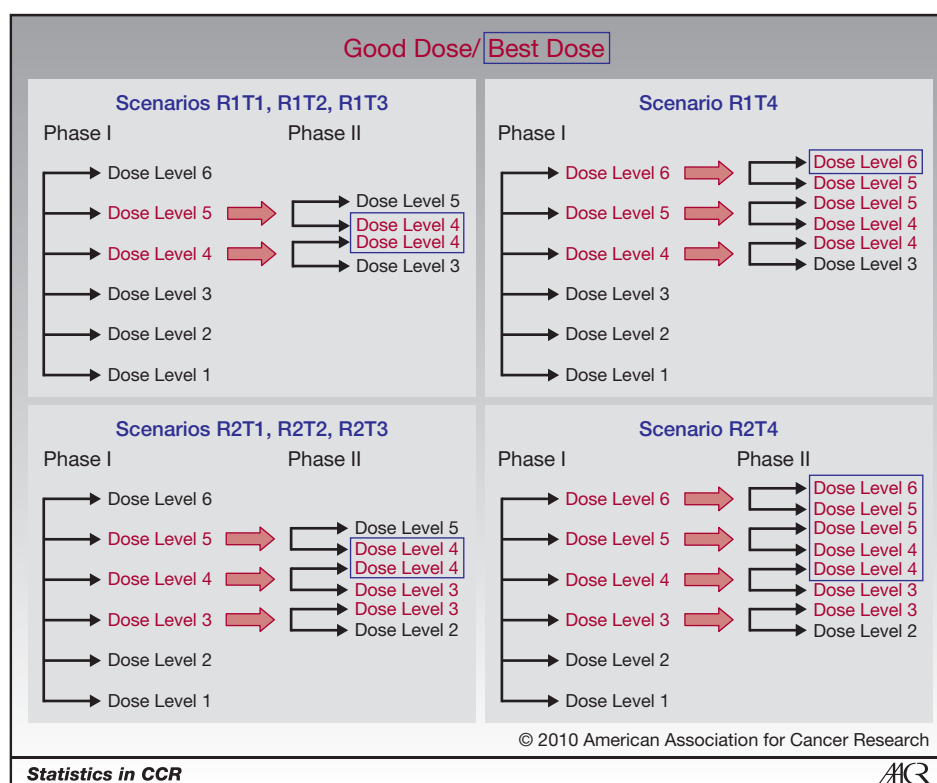
Figure 3. Possible ways to select a good dose or the best dose level using our proposed design for response scenarios R1 and R2. Good dose levels as defined by toxicity and efficacy are in red. The best dose levels are circled in blue.

Table 3. Overall probability of selecting the *best* or *a good* dose level

Efficacy Scenario	Toxicity Scenario	Best Dose Level	Good Dose Level	Probability of Picking *Best* Dose with Our Proposed Design	Probability of Picking a *Good* Dose with Our Proposed Design	Probability of Picking *Best* Dose with Traditional Phase I-II Design	Probability of Picking *Good* Dose with Traditional Phase I-II Design
R1	T1	4	4	0.16	0.16	0.21	0.21
R1	T2	4	4	0.15	0.15	0.18	0.18
R1	T3	4	4	0.29	0.29	0.30	0.30
R1	T4	6	4–6	0.61	0.66	0.84	0.89
R2	T1	4	3, 4	0.46	0.55	0.21	0.51
R2	T2	4	3, 4	0.43	0.43	0.19	0.46
R2	T3	4	3, 4	0.53	0.58	0.31	0.48
R2	T4	4–6	3–6	0.89	0.93	0.89	0.93
R3	T1	1–4	1–4	0.93	0.93	0.86	0.86
R3	T2	1–4	1–4	0.90	0.90	0.80	0.80
R3	T3	1–4	1–4	0.85	0.85	0.78	0.78
R3	T4	1–6	1–6	0.97	0.97	0.97	0.97

Our simulations of the phase I part of the trial are also summarized in Table 1. We chose a high correlation or odds ratio between efficacy and toxicity for simulating the efficacy and toxicity data. The log odds ratio we chose for all our simulation studies is 4.6. The dose levels marked by footnote "a" indicate the probability for correctly reaching the MTD. The 3+3 design is very conservative. The probability of reaching the level above the MTD is, in general, small. In the scenarios with dose level 4 being the MTD, the probability of correctly identifying the MTD or the dose level below is similar and, in general, somewhere between 20 and 30%. Scenarios T2 and T3 both assume a monotone increase in toxicity with dose level 4 being the MTD. The only difference is that scenario T2 has a steeper increase than scenario T3; dose level 5 for T2 is set at 40% well above the MTD, whereas dose level 5 for T3 is set at 35% just slightly above the MTD. As a result, the mass of the probability distribution for T3 is moved to the right compared with T2, and the probability of correctly reaching the MTD (level 4) or the level above the MTD (level 5) is higher than for T2.

Fig. 3 illustrates the possible ways to reach a good dose or the best dose for response scenarios R1 and R2. The schema for scenarios R1T1, R1T2, and R1T3 on the top left side of Fig. 3 illustrates the 2 possible ways to reach dose level 4 (the only level with acceptable toxicity and efficacy). If the MTD is correctly identified in phase I, the patients will be randomized between dose levels 4 and 3, and ending up with the best dose level is a possibility. If the phase I trial picks level 5 (the dose level right above the MTD), patients will be randomized between level 5 and 4 in the phase II portion of the trial, and again, selecting the best dose level is a possibility. The more "best" dose levels there are, the more ways to correctly identify the best dose can be found. For scenario R3T4 (not shown), all dose levels are considered "best" levels.

Table 2 summarizes our phase II simulation results for the various efficacy and toxicity scenarios for the case in which the phase I trial correctly identified the MTD. In this case, patients are being randomized to 2 dose levels: the MTD (arm 1) and the dose level immediately below the MTD (arm 2). Similar simulation studies were done for the case in which the phase I trial identifies the dose level above or below the MTD as the correct dose. In addition to evaluating the power, the probability of more than 33% of patients experiencing an MTD in each arm was determined. Finally, the probability of selecting the better dose level (in our example arm 1) by using our modified selection design taking into account both efficacy and toxicity is summarized in the last 2 columns of this table.

Our simulations of a traditional single-arm phase II trial with 40 patients, if the MTD is correctly determined in the phase I trial, revealed a much larger power when comparing the results with Table 2 as expected, owing to larger sample size.

Specifically, Table 3 compares our results to the traditional sequence of a phase I trial followed by a single-arm phase II trial at the dose level determined by the phase I trial, using the same total sample size. We have to keep in mind that, due to the discreteness of the binomial distribution, the alpha levels that are determined by the efficacy boundary in the 2 examples (single arm with 40 patients versus 2 arm with 20 patients each) are not identical. The levels are 0.05 for the traditional single-arm trial and 0.07 for each of the arms in the phase II selection design.

In general, the probability of picking a good or best dose is similar or higher in our proposed design than in the traditional design. The traditional design fares better if the true efficacy is close to the alternative hypothesis for the MTD (scenario R1). In that case, doubling the sample size in the phase II portion yields a considerably higher power and, thus, a higher probability of the phase II portion of having a positive outcome. If the underlying toxicity is not uniformly low relative to the maximum toxicity cut off (i.e., excluding T4 in our examples), the

difference in probability of picking a good or best dose is, at most, 5% higher than in our proposed design. The only scenario in which the traditional design fares considerably better is the scenario in which the true efficacy is close to the alternative hypothesis for the MTD (R1) and toxicity is negligible (T4). In all other response scenarios, our proposed design does better than the traditional design. This better performance is particularly true for the efficacy scenario R2 and the toxicity scenarios T1 to T3 in which toxicity is not negligible. In those scenarios, finding the best dose is about twice as likely as in the traditional design.

We also explored the scenario with toxicity increasing with dose level, but staying below 33%. More specifically, we added a toxicity scenario T5 for which toxicity increases linearly from 0.05 for dose level 1 up to 0.25 for dose level 6 (not shown in figures or tables). We examined the properties of this toxicity scenario in combination with the response scenario R2 (linear increase up to dose level 4, then leveling off). In this case, the MTD is level 6 and we have 3 "best" dose levels (4, 5, and 6) and 4 "good" dose levels (3, 4, 5, and 6). We found that the performance of that scenario was similar to the performance of R2T4 (T4, toxicity at 0.05 for all levels), but the probabilities for picking the best and a good dose level were lower than for R2T4 as toxicity was finite. More specifically, the probability for picking a best dose is 0.55 for our proposed design and 0.56 for the traditional phase I-II design; whereas the probability for picking a good dose is 0.73 for both our proposed and the traditional phase I-II design. We also investigated a 4th response scenario R4, with the response rate monotonically increasing to more modest levels (40% maximum). The power for $N = 20$ is considerably lower than for $N = 40$ in this case, and the traditional design does better.

The 3+3 phase I trial design is designed to be very conservative (see Table 2). For the toxicity scenarios T1 and T2, the probability of determining the dose level above the MTD as the correct dose level is 8% or lower. Thus, it is very unlikely for efficacy scenarios R1 and R2 to eventually arrive at the best dose level by reaching the dose level above the MTD first. On the other hand, the probability of reaching the dose level right below the MTD as the correct dose level in a phase I trial is often as high as that of reaching the MTD. A possible consideration that would greatly increase the probability of reaching the best dose level with our seamless phase I-II trial design would be to randomize patients to 3 dose levels: the dose level determined to be the MTD by the phase I trial, and the dose levels right below and right above that dose. This use of the trial design would be a possibility if there was strong evidence in animal studies and the understanding of the pathways of activity that this new agent was not toxic. This implementation of the trial design would obviously require continuous toxicity monitoring in the phase II portion and appropriate toxicity stopping rules for the higher dose levels. Three dose levels would also allow simple logistic regression modeling under the assumption of smooth dose and toxicity profiles to reduce variance of estimates of response or toxicity at a given toxicity level.

In summary, the design proposed in this manuscript does better in most cases than a traditional design using the same overall sample size. We chose a 3+3 trial design for the simulation studies of the phase I portion of the trial; other phase I trial designs such as a CRM or accelerated titration could have been used instead. A possibility would be to slightly increase the sample size in the phase II portion of the proposed design and, thus, assuring the highest probability of finding the most efficacious and least toxic dose level. This design allows assessing a few dose levels more closely for both efficacy and toxicity and greater certainty of having correctly determined the best dose level before launching into a large efficacy trial. It should, thus, be considered even in the scenarios in which a slightly larger sample size may be required.

References

1. Storer BE. Choosing a phase I design. In:Crowley J, Power Ankerst D, editors. Handbook of statistics in clinical oncology. 2nd ed. Boca Raton (FL): CRC Press; 2006. p. 59–75.
2. Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade offs. Biometrics 2004;60:684–93.
3. Zhang W, Sargent DJ, Mandrekar S. An adaptive dose-finding design incorporating both toxicity and efficacy. Stat Med 2006;25:2365–83.
4. Hunsberger S, Rubinstein LV, Dancey J, Korn EL, et al. Dose escalation trial designs based on a molecularly targeted endpoint. Stat Med 2005;24:2171–81.
5. Jain RK, Lee JJ, Hong D, Markman M, Gong J, Naing A, et al. Phase I oncology studies: evidence that in the era of targeted therapies patients on lower doses do not fare worse. Clin Cancer Res 2010;16:1289–97.
6. Hanrahan EO, Heymach JV. Vascular endothelial growth factor receptor tyrosine kinase inhibitors vandetanib (ZD6474) and AZD2171 in lung cancer. Clin Cancer Res 2007;13: 4617s.
7. Liu PY, Moon J, LeBlanc M. Phase II selection designs. Chapter 9. Handbook of statistics in clinical oncology. 2nd ed. Boca Raton (FL): CRC Press; 2006.
8. Yee TW. The VGAM R-Package: vector generalized linear and additive models. 2009.
9. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. London: Chapman & Hall; 1989.
10. Dale JR. Global cross-ratio models for bivariate, discrete, ordered responses. Biometrics 1986;42:909–17.