Accelerated Titration Designs for Phase I Clinical Trials in Oncology

Richard Simon, Boris Freidlin, Larry Rubinstein, Susan G. Arbuck, Jerry Collins, Michaele C. Christian*

Background: Many cancer patients in phase I clinical trials are treated at doses of chemotherapeutic agents that are below the biologically active level, thus reducing their chances for therapeutic benefit. Current phase I trials often take a long time to complete and provide little information about interpatient variability or cumulative toxicity. *Purpose:* Our objective was to develop alternative designs for phase I trials so that fewer patients are treated at subtherapeutic dose levels, trials are of reduced duration, and important information (i.e., cumulative toxicity and maximum tolerated dose) needed to plan phase II trials is obtained. *Methods:* We fit a stochastic model to data from 20 phase I trials involving the study of nine different drugs. We then simulated new data from the model with the parameters estimated from the actual trials and evaluated the performance of alternative phase I designs on this simulated data. Four designs were evaluated. Design 1 was a conventional design (similar to the commonly used modified Fibonacci method) using cohorts of three to six patients, with 40% dose-step increments and no intrapatient dose escalation. Designs 2 through 4 included only one patient per cohort until one patient experienced dose-limiting toxic effects or two patients experienced grade 2 toxic effects (during their first course of treatment for designs 2 and 3 or during any course of treatment for design 4). Designs 3 and 4 used 100% dose steps during this initial accelerated phase. After the initial accelerated phase, designs 2 through 4 resorted to standard cohorts of three to six patients, with 40% dose-step increments. Designs 2 through 4 used intrapatient dose escalation if the worst toxicity is grade 0-1 in the previous course for that patient. *Results:* Only three of the actual trials demonstrated cumulative toxic effects of the chemotherapeutic agents in patients. The average number of pa-

tients required for a phase I trial was reduced from 39.9 for design 1 to 24.4, 20.7, and 21.2 for designs 2, 3, and 4, respectively. The average number of patients who would be expected to have grade 0-1 toxicity as their worst toxicity over three cycles of treatment is 23.3 for design 1, but only 7.9, 3.9, and 4.8 for designs 2, 3, and 4, respectively. The average number of patients with grade 3 toxicity as their worst toxicity increases from 5.5 for design 1 to 6.2, 6.8, and 6.2 for designs 2, 3, and 4, respectively. The average number of patients with grade 4 toxicity as their worst toxicity increases from 1.9 for design 1 to 3.0, 4.3, and 3.2 for designs 2, 3, and 4, respectively. *Conclusion:* Accelerated titration (i.e., rapid intrapatient drug dose escalation) designs appear to effectively reduce the number of patients who are undertreated, speed the completion of phase I trials, and provide a substantial increase in the information obtained. [J Natl Cancer Inst 1997;89:1138-47]

There has been considerable recent interest in new designs for phase I clinical trials. With currently used designs, many patients are treated at doses below the biologically active level, minimizing the opportunity for antitumor response (*1*). Although most patients who participate in phase I trials hope to obtain

*Affiliations of authors: R. Simon, L. Rubinstein, S. G. Arbuck, M. C. Christian, Cancer Therapy Evaluation Program, Division of Cancer Treatment, Diagnosis, and Centers, National Cancer Institute, Bethesda, MD; B. Freidlin, The Emmes Corporation, Potomac, MD; J. Collins, U.S. Food and Drug Administration, Rockville, MD.

Correspondence to: Richard Simon, D.Sc., National Institutes of Health, Executive Plaza North, Rm. 739, Bethesda, MD 20892.

See ''Notes'' following ''References.''

therapeutic benefit from promising new experimental treatments, few achieve this objective (*2*). Whereas most patients would not have derived benefit from drugs studied in phase I trials, even if treated at the maximum tolerated dose (MTD), treating patients at substantially lower doses is likely to further reduce whatever chance for benefit might exist.

A second problem with current designs is that phase I trials may take a long time to complete, especially when the starting dose is far below the MTD (*3*). Current phase I trials also provide almost no information about variability among patients in the dose that can be tolerated without dose-limiting toxicity (DLT) or about whether there is evidence of cumulative toxicity.

In phase I trials of new drugs, the starting dose is usually one tenth of the LD_{10} (i.e., the dose that is lethal to 10% of animals) in the most sensitive animal species in which toxicology studies have been performed. Dose steps are defined by a modified Fibonacci series in which the increments of dose for succeeding levels are 100%, 67%, 50%, and 40%, followed by 33% for all subsequent levels. Three patients are usually treated at a dose level and observed for acute toxicity for one course of treatment before any more patients are entered. If none of the three patients experience DLT, then the next cohort of three patients is treated at the next higher dose. If two or more of the three patients experience DLT, then three more patients are treated at the next lower dose unless six patients have already been treated at that dose. If one of three patients treated at a dose experiences DLT, then three more patients are treated at that same level. If the incidence of DLT among those six patients is one in six, then the next cohort is treated at the next higher dose. In general, if two or more of the six patients treated at a dose level experience DLT, then the MTD is considered to have been exceeded, and three more patients are treated at the next lower dose as described above. The MTD is defined as the highest dose studied for which the incidence of DLT was less than 33%. Usually dose escalation for subsequent courses in the same patient, intrapatient dose escalation, is not permitted.

In this article we will describe alternative phase I designs that attempt to overcome some of the problems described above. We will then report the results of a computer simulation study conducted to evaluate the performance of alternative designs. The designs will be evaluated with regard to safety, the extent to which they provide patients the opportunity to be treated at higher doses more likely to provide antitumor response, the number of patients and time required to complete the trial, and the amount of information obtained.

Several alternative approaches to the design of phase I trials have been discussed in previous years. Collins et al. (*3*) recommended accelerating the dose escalations in humans by using the plasma drug $C \times T$ (i.e., the area under the concentration versus time curve) value at the LD_{10} in the mouse as the target exposure. This provides a pharmacokinetic basis for dose escalation, but is limited to clinical situations where a sensitive assay for the active drug moieties is available and where interspecies pharmacodynamic differences do not exist for the drug.

Storer (*4*) introduced the concept that the objective of a phase I trial is to estimate the dose that causes DLT in a specified proportion (e.g., 25% of the patients), and that this MTD should be estimated by fitting a logistic model to the dose versus DLT data. Storer also proposed using a single patient per dose level until the first DLT is observed.

Several investigators (*5-10*) have considered Bayesian designs. This approach makes use of a model relating dose administered to the probability of DLT. The parameters of the model are unknown initially, but some prior probability distribution for their values is assumed to be available based on preclinical data or experience with other drugs. As patients are treated, the probability estimates of the unknown parameters are updated based on the actual toxicity experience observed. Each patient is assigned the dose predicted to result in DLT for a target percentage (e.g., 25%) of the population.

Mick and Ratain (*11*) used a linear model relating white blood cell (WBC) count nadir to dose and pretreatment WBC count. They sequentially estimated the regression parameters of the model as data accumulated and individualized the dose based on pretreatment WBC count in an attempt to achieve a specified optimal WBC count nadir. Their approach predicts the optimal dose for each patient is based on pretreatment patient characteristics.

None of the designs described above considers how patients should be treated after the first course, nor do they use information obtained from subsequent courses. Except for the approach of Mick and Ratain (*11*), they do not consider interpatient variability or use information about toxic effects less than DLT.

Sheiner et al. (*12,13*) have argued for the use of titration (or intrapatient dose-escalation designs) for evaluating drug efficacy for diseases where the condition of a patient remains stable over a period of time. Titration designs involve dose escalation within patients until the desired biologic effect is obtained. If analyzed properly, they can provide information about interpatient variability in dose–response effects. The analysis of titration designs has been studied (*14,15*), but this approach has not been discussed in the context of phase I trials in oncology.

Methods

Phase I Designs Studied

The designs we evaluated differ with regard to the escalation/de-escalation rules for the first-course treatment of subsequent cohorts of patients as indicated in the ''Appendix'' section. Design 1 is the standard design described above. The other dose-escalation methods are based on a four-grade scale for defining the highest level of overall toxicity during each course of therapy. This scale can be defined differently to accommodate different clinical situations. For the purposes of this article, we have related the toxicity experience to grading scales commonly used in oncology, such as the National Cancer Institute Common Toxicity Criteria, and have described the levels as follows: none-mild (grades 0-1), moderate (grade 2), dose limiting (grade 3), and unacceptable (grades 4-5). Consistent with recent practice, we have not considered grade 3 neutropenia unaccompanied by either fever or infection to be dose limiting. We have grouped no toxicity with grade 1 toxicity because of the difficulty of determining whether mild abnormalities are drug or illness related in patients with cancer.

Design 2 treats one patient per dose level until one patient exhibits DLT or two patients exhibit grade 2 toxicity during their first course of treatment. At that time, the escalation plan switches to design 1. That is, two additional patients are accrued at the dose that triggered the switch, and three to six patients are treated in that and each subsequent cohort. This approach offers the possibility of speeding up the trial and reducing the number of patients assigned to low doses. It uses the first instance of first-course DLT to trigger the switch as proposed by Storer (*4*). It also uses first-course grade 2 toxicity to provide an added element of caution. We use the second instance of grade 2 toxicity for practical reasons, since it is often difficult to determine whether a grade 2 toxicity is drug related in a heterogeneous population of very ill patients.

Designs 3 and 4 also use only one patient per cohort during the early stage of the trial, but they incorporate more rapid dose escalation by using double-dose steps during this stage. With design 3, the single-patient-cohort stage of the trial also terminates when one patient experiences first-course DLT or two patients experience first-course grade 2 toxicity. With design 4, this accelerated stage terminates when the first instance of DLT or the second instance of grade 2 toxicity is observed in any course of treatment. In either case, after the rapid escalation stage terminates, subsequent cohort sizes are three to six patients and single-dose escalation steps are used as in design 1.

The Appendix also describes two approaches to individualizing dose through intrapatient dose modification. Intrapatient modification option A is the one most commonly used. There is no intrapatient dose escalation, only de-escalation. If the toxicity is dose limiting or worse in a course of chemotherapy, then the dose is reduced one level for the next course. Otherwise, the dose stays the same for the next course. Intrapatient modification option B permits escalation for each patient if the toxicity is grade 0-1 in the previous course for that patient. If the toxicity is moderate (grade 2), the dose remains unchanged. However, if the toxicity is DLT or worse, the dose is reduced. Designs 3 and 4 use two-dose-step (100%) intrapatient escalations during the initial accelerated phase of the trial, although de-escalations are always by single-dose steps. We have combined the standard cohort escalation design with the standard intrapatient dose modification option (A) as design 1 and have combined accelerated cohort escalations with the intrapatient dose escalation option (B) as designs 2 through 4. We will also provide results, however, for the mixed designs such as escalation option A with designs 2 through 4.

The accelerated designs are intended for use in phase I trials of drugs that have not been used previously in humans, where only preclinical information will be available for selecting a starting dose. Starting doses in these cases are often quite low, and designs that limit the number of patients treated at very low doses may be particularly useful.

Methodologic Approach

To evaluate alternative phase I designs, we wished to use data from actual phase I trials as much as possible. This could not be done directly because past trials were conducted with a particular escalation plan and we wished to evaluate new plans. Instead, we fit a stochastic model to data from past phase I trials. We then simulated new data from the model with the parameters estimated from the actual trials and evaluated the performance of alternative escalation designs on these simulated data. For any particular phase I trial, we generated 1000 simulated sets of data to reliably estimate the relative performance of the alternative designs. We repeated this for 20 different actual phase I trials of nine different drugs.

We required that the model we used be able to represent different levels of worst toxicity, not just presence or absence of DLT, and that the toxicity level experienced in a particular course would be determined by the dose administered in that course and the total dose administered in previous courses. We required that both interpatient and intrapatient variability be represented. We used the following model. Suppose that the ith patient receives dose d_{ij} during course j and has received a total dose of D_{ij} for courses previous to j. We let the coefficient α represent the influence of prior total dose ($\alpha = 0$ indicates no cumulative toxicity) and let the magnitude of toxicity increase logarithmically with dose. We introduced a random number β_i, normally distributed with mean μ_β and variance σ_β^2. This variable represents the interpatient variability in sensitivity to the toxic effects of the drug. We also introduced a random number ε_{ij} normally distributed with mean zero and variance σ_ε^2, to represent the intrapatient variability in toxic response for a given patient receiving a given dose. These terms and random variables determine the magnitude of worst toxicity represented by

$$y_{ij} = \log(d_{ij} + \alpha D_{ij}) + \beta_i + \varepsilon_{ij}. \qquad [1]$$

If this value y_{ij} was less than a specified constant K_1, then patient i was considered to have experienced less than grade 2 toxicity during course j with dose d_{ij}. If the value of y_{ij} was greater than K_1 but less than K_2, then the toxicity level was taken to be grade 2; if the value was greater than K_2 but less than K_3, then the toxicity was considered to be dose limiting; and if y_{ij} was greater than K_3, then the toxicity was considered unacceptable. The values of the random numbers β_i vary across patients, but the same β_i was used for all treatment courses of the ith patient, while the within patient variability values ε_{ij} change from patient to patient as well as across courses.

This model can be viewed as a generalization of the K_{max} model used by Sheiner et al. (12,13). The above expression is equivalent to

$$\frac{e^{y_{ij}}}{1 + e^{y_{ij}}} = \frac{(d_{ij} + \alpha D_{ij})}{e^{-(\beta_i + \varepsilon_{ij})} + (d_{ij} + \alpha D_{ij})}. \qquad [2]$$

The right-hand side of this equation is similar to the K_{max} model. The stimulus is of the form $d_{ij} + \alpha D_{ij}$ and the level giving 50% maximum response (exclusive of cumulative toxicity) is taken as a random variable, with mean approximately $e^{-\mu_\beta}$ and with a component identified with interpatient variability and a component associated with intrapatient variability. Our model measures toxicity in a categorical rather than continuous manner. Since the scale of the constants K_1, K_2, and K_3 is arbitrary, the fact that the left side of the equation involves a transformation of the originally defined y_{ij} does not matter. In fact, it can be shown that our model can be viewed as a generalization of the model of Chou and Talalay (16) in which the stimulus $d_{ij} + \alpha D_{ij}$ and 50% value are raised to a power p. With a categorical response in which the K's may be fit from the data, however, the power p is not identifiable, and the model is equivalent to that shown in equation 1.

The value σ_ε^2 represents the amount of intrapatient variability unexplained by current and previous doses. Setting $\sigma_\varepsilon^2 = 0$ means that the toxicity experienced by a patient is determined entirely by the doses and by patient characteristics that do not change from day to day. The value of σ_β^2 represents the amount of interpatient variability. Setting $\sigma_\beta^2 = 0$ means that patients entered in the clinical trial do not differ in their ability to tolerate the drug under study.

For these simulations, we used 40% increments between dose levels. With 40% increments, two-dose levels represent approximately a doubling of the dose because $1.4^2 = 1.96$. A 40% increment is close to the 33% increment that is used after the first few dose levels of trials based on the modified Fibonacci approach with which phase I investigators are familiar. Because interpatient variability in patient pharmacokinetic parameters and intrapatient variability in day-to-day susceptibility to toxicity are often substantial, it is usually not realistic to expect that one can estimate more precisely than to within 40% the dose that will give a desired level of biologic effect (17).

For all the simulations, we used $\mu_\beta = 0$, although the results are independent of this parameter. Table 1 shows the maximum likelihood estimates of the model parameters for the 20 actual phase I clinical trials studied. These trials were selected for a related study of nonstandard dose-escalation procedures. Although they were selected initially because they were planned to use nonstandard dose-escalation methods, only 9.5% of the patients received intrapatient dose escalation. Detailed information about the characteristics of these trials will be addressed in a separate report.

Only three of the 20 trials showed any evidence of substantial cumulative toxic effects as seen from the column labeled α in Table 1. Two of these studies involved the drug pyrazine diazohydroxide (PZDH) administered as a bolus every 3 weeks initially, but the interval between courses was eventually lengthened to 4-6 weeks because of delayed recovery from myelosuppression. Trial T90-156 administered PZDH daily for 5 days every 4-6 weeks, and no evidence of cumulative toxicity was obtained from our model parameters for that trial. The third trial showing evidence of cumulative toxicity involved flavone acetic acid (FAA). This latter trial was the only phase I trial with FAA that demonstrated cumulative toxicity. It differed from the other four FAA trials in that it used a weekly schedule of administration.

The standard deviations for intrapatient (σ_ε) and interpatient (σ_β) variability varied substantially. The larger values of σ_ε seen could represent true biologic variability or may reflect the difficulty of distinguishing drug-related toxicity from manifestations of illness for very sick patients. We used the original treating physician's assessment as to whether toxicity was drug related. Many of these patients were taking concomitant medications (not anticancer drugs) that may have influenced the toxicity experienced, and, in some cases, there may also have been nonstandardized treatment delays as a result of previous toxicity. With prospective use of titration designs, we expect that there will be more attention to these issues than could be the case in a retrospective analysis of a database.

The K_1 value is given in terms of (K_1 – log starting dose)/log 1.4 because this value represents approximately the number of 40% dose steps between the starting dose and the dose at which the average patient has a 50% chance of experiencing grade 2 or worse toxicity (since $\mu_\beta = 0$). The distance between other K values is similarly presented. Seven of the actual trials did not have any patients who experienced grade 4 toxicity. For these cases, the estimate of K_4 is very large by default, but the specific value is not meaningful.

Table 1. Estimates of model parameters for 20 phase I clinical trials

Drug	Trial	α	$(K_1 - \ln d_0)/\ln 1.4$*	$(K_2 - K_1)/\ln 1.4$	$(K_3 - K_2)/\ln 1.4$	σ_β	σ_ϵ
Flavone acetic acid	85-168	0	16.2	6.9	35†	.26	1.9
Flavone acetic acid	85-244	0	16.1	8.4	29†	2.9	.85
Flavone acetic acid	86-004	0	4.4	2.4	0.95	.47	.59
Flavone acetic acid	86-017	.24	8.0	2.9	2.2	0	.83
Flavone acetic acid	86-060	0	18.5	6.4	20†	.006	2.8
Piroxantrone	86-227	.08	8.4	2.7	2.3	1.03	.42
Piroxantrone	86-268	0	16.4	13.3‡	9.5‡	0	1.8
Chloroquinoxaline sulfonamide	88-114	.04	17.3	2.6	1.6	.88	.87
Chloroquinoxaline sulfonamide	88-127	0	13.7	4.6	2.9	.62	.90
Pyrazine diazohydroxide	89-053	.56	6.7	1.3	2.0	.37	.50
Pyrazine diazohydroxide	89-175	.24	6.6	1.3	0.53	.002	.65
Pyrazine diazohydroxide	90-156	.02	4.6	.53	.56	.001	.18
Pyrazoloacridine	90-073	.04	8.9	1.0	1.3	.24	.32
Cyclopentenylcytosine	91-018	0	4.4	.83	0.18	.19	.26
Fostriecin	91-106	.04	3.5	3.6	4.5	1.06	.54
Fostriecin	91-196	0	6.3	7.2	18†	.58	1.6
9-Aminocamptothecin	92-108	0	6.4	.48	0.39	.24	.11
9-Aminocamptothecin	92-186	0	6.0	.51	1.1	.35	.27
Penclomedine	93-087	.05	6.0	3.7	15†	.68	.81
Penclomedine	93-125	0	5.8	2.0	17†	.43	.53

*d_0 = starting dose.
†No grade 4 toxicity.
‡No grade 3+ toxicity.

It may be noted in Table 1 that the parameter estimates for different trials of the same drug sometimes vary substantially. This is due to a variety of causes, but the estimates provide a wide range of conditions for generating simulated data with which to compare alternative escalation designs.

Results

Comparison of Designs

The distribution of the highest dose level at which fewer than two instances of DLT occurred was very similar for the four designs for all of the 20 sets of parameters studied (Fig. 1). The true MTD was defined as the largest dose level for which the probability of first-course DLT or worse was less than .25, computed from the model using each set of parameters in Table 1. For simulations with each set of parameters, we tabulated the accuracy of the highest dose level with fewer than two instances of first-course DLT as a predictor of the true MTD. Fig. 1 shows that the four designs performed similarly in this regard. Although designs 2 through 4 use many fewer patients than design 1, in the dose range of interest, they have similar sample sizes. As will be seen later, fitting the model to data from a phase I trial provides a much richer set of information with which to plan phase II development. Fig. 1 demonstrates, however, that even with regard to the traditional estimate of phase II dose, accuracy is not sacrificed by the accelerated designs.

Fig. 2 shows histograms of the average number of patients required in the simulated trials for each design. In each graph, the sum of the heights of the bars is 20, the number of sets of parameters that is simulated. The *x* axis represents the average number of patients accrued in the 1000 simulated trials with each of the 20 sets of parameters. The standard design has a very broad distribution of sample size. For six of the sets of parameters, design 1 required more than 55 patients. For the 20 sets of parameters, design 1 required an average of 39.9 patients (median, 36.7 patients). Design 1 required substantially more pa-

tients than did the other designs. Design 2, which uses single 40% dose steps, required an average of only 24.4 patients (median, 21.8 patients). As seen in Fig. 2, the distribution of the number of patients is much narrower for design 2 than for design 1. Designs 3 and 4 also compare very favorably with design 1 with mean numbers of patients of 20.7 and 21.2, respectively, and median numbers of patients of 19.3 and 19.1, respectively. Design 2 does not require many more patients than the accelerated designs that use double dose steps. As will be seen below,
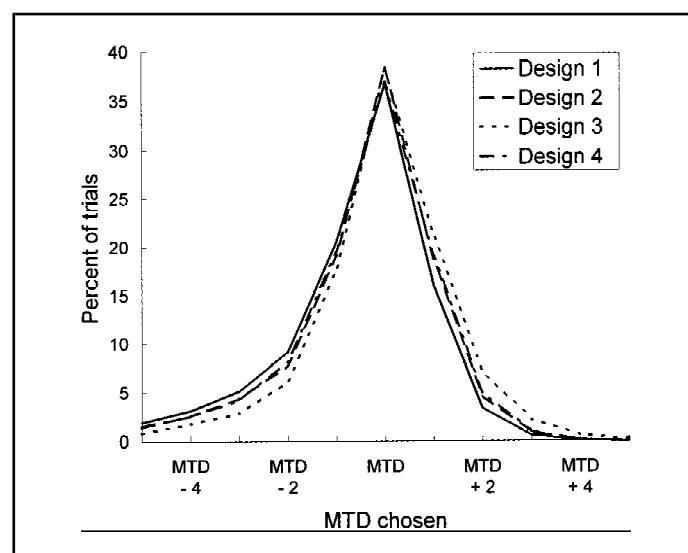
Fig. 1. Distribution of the maximum tolerated dose (MTD) chosen in simulated phase I trials averaged over the 20 sets of model parameters and 1000 replications for each set of parameters. In simulations, MTD is chosen in the traditional way as the largest dose level at which six patients are started and fewer than two experience first-course dose-limiting toxicity (DLT) or worse. True MTD is defined as the highest dose level for which the probability of first-course DLT or worse is less than .25, computed from the model with the use of each set of parameters in Table 1.

Fig. 2. Histograms of the average numbers of patients for simulated phase I trials with each of the 20 sets of model parameters. Averages are based on 1000 replications for each set of parameters. Total height of bars in each panel equals 20, the number of sets of model parameters.
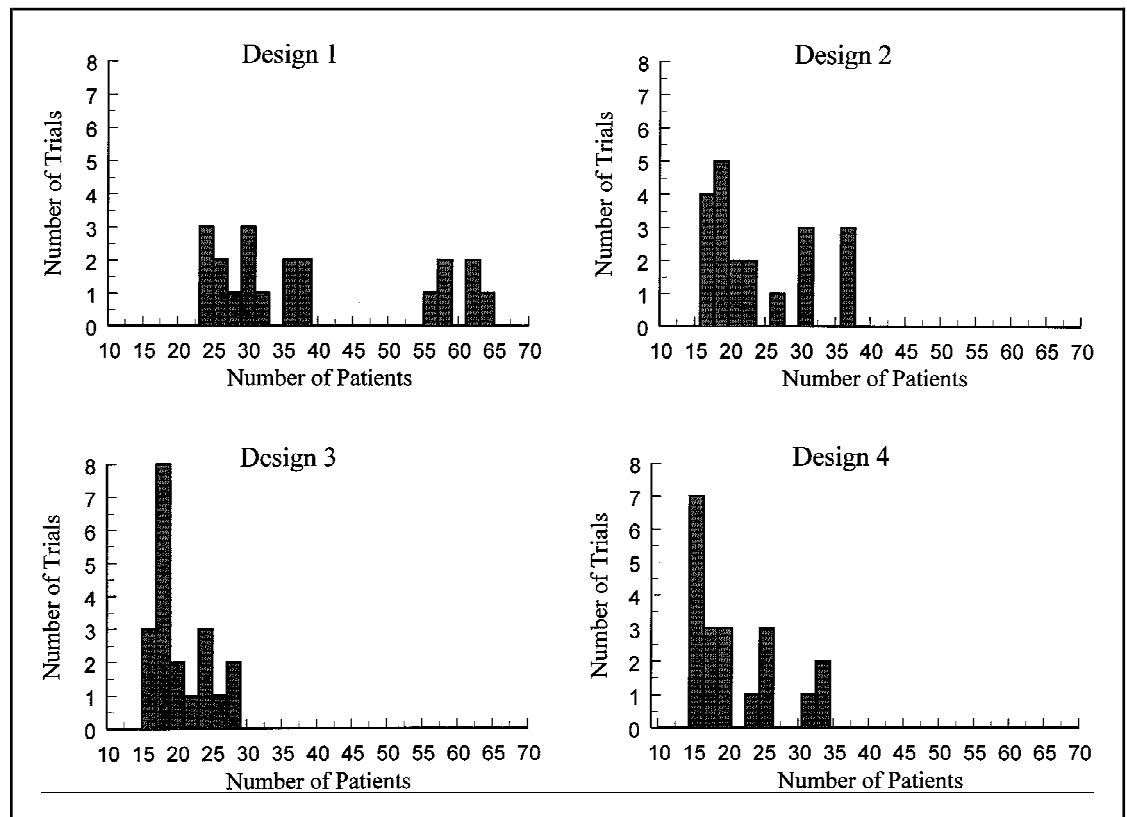
the important difference between design 1 and the others is largely due to a reduction in patients treated early at subtherapeutic doses, where designs 2 through 4 accrue only one patient per level.

Another question of some importance is whether a reduction in the number of patients translates into a reduction in the duration of the trial. When eligible patients are very limited, the number of patients is closely associated with the duration of the trial. But if eligible patients are readily available, then it would take little more time to place three patients on a dose level than to place a single patient. Therefore, we also tabulated the number of cohorts required for each design, as shown in Fig. 3. The advantage of design 2 over design 1 with regard to the average number of patients does not translate into an advantage in the number of cohorts required. In fact, design 1 requires slightly fewer cohorts because design 2 sometimes overshoots its target and requires more cohorts at de-escalated levels. Designs 3 and 4, however, show substantial savings over designs 1 and 2 because of their use of double dose steps during the initial stage of the trials.

Fig. 4 shows the toxicity experience in the application of these designs to the phase I trials. In these simulations, we have assumed that all patients stay in the study for three courses of treatment and have tabulated the distribution of worst toxicity over these courses for each patient. For each set of parameters and each design, we have calculated the average number of patients whose worst toxicity was grade 0-1, grade 2, grade 3, or grade 4. This average was computed based on 1000 simulations for each of the 20 sets of design parameters. With the standard design 1, the average number of patients who have grades 0-1 toxicity as their worst toxicity over three cycles of treatment is

23.3. This number is substantially reduced for all of the newer designs; 7.9 for design 2, 3.9 for design 3, and 4.8 for design 4. Therefore, the number of undertreated patients is substantially reduced. This reduction is achieved with some increase in the number of patients with worst toxicity grade 3 or 4. The average number of patients with worst toxicity grade 3 increases from 5.5 with design 1 to 6.2, 6.8, and 6.2 for designs 2, 3, and 4, respectively.

Fig. 4 shows that the average number of patients with grade 4 toxicity increased from 1.9 with design 1 to 3.0, 4.3, and 3.2 for designs 2, 3, and 4, respectively. Hence, in comparing design 2 to design 1, on average, there is a reduction of about 15 patients per trial whose highest level of toxicity is grade 0-1 and an average increase of 1.8 patients per trial whose highest level of toxicity is grade 3-4. Design 4 provides a reduction of about 18 undertreated patients per trial and an average increase of about 2.1 overtreated patients. Design 3 provides a reduction of about 19 undertreated patients per trial, for an average increase of 3.7 overtreated patients. Hence, design 3 appears to have no real advantage over design 4. Although the average number of patients with worst toxicity grade 3-4 is not substantially increased using designs 2 through 4 compared with design 1, the proportion of patients with grades 3-4 toxicity is substantially increased. This is because designs 2 through 4 substantially reduce the expected number of patients with worst toxicity grade 0-1 and the total number of patients on trial compared with design 1. With design 1, a weighted average (taken over the 20 parameter sets, weighted by average sample size) of about 18% of patients experience grade 3-4 toxicity during some course of treatment. For designs 2, 3, and 4, the percentages are about 38%, 53%, and 45%, respectively. For grade 4 toxicity alone, the

Fig. 3. Histograms of the average number of cohorts of patients for simulated phase I trials with each of the 20 sets of model parameters. Averages are based on 1000 replications for each set of parameters. Total height of bars in each panel equals 20, the number of sets of model parameters. Number of cohorts reflects time to completion when there is an excess of patients available for entry in the trial.
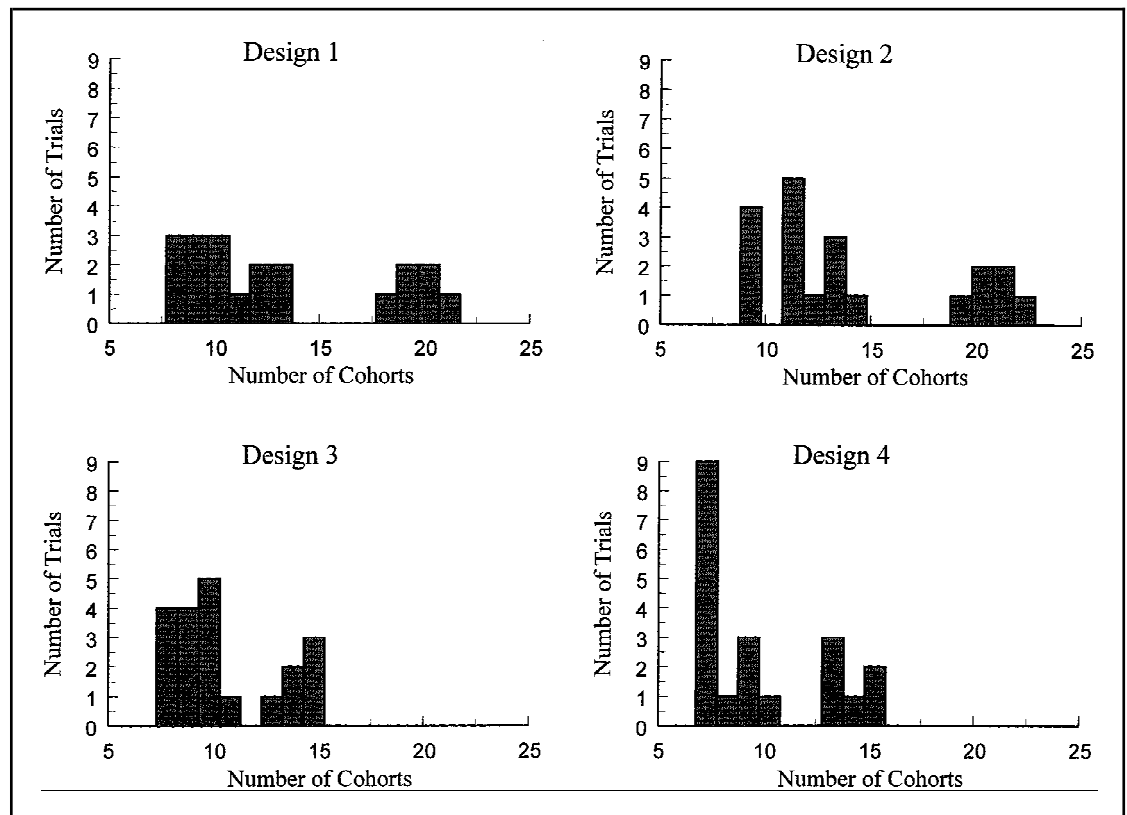
percentages are 5%, 12%, 20%, and 15% for designs 1 through 4, respectively.

There are six sets of parameters by use of design 1 for which three or more patients are expected to experience grade 4 toxicity. The trial with the largest number of such patients was T89-175. This is a PZDH trial with $\alpha = .24$. The PZDH trial with $\alpha = .56$ (T89-053) and the FAA trial with $\alpha = .24$ are also included in this set of six trials. It is not surprising that trials with a substantial amount of cumulative toxicity should result in patients experiencing grade 4 toxicity, even without using intrapa-

Fig. 4. Expected average number of patients in each toxicity grade for the 20 phase I trials studied. All patients are assumed to stay in the trial for three courses of therapy. Toxicity grade is the highest toxicity level experienced over the three courses.

tient dose escalation. The other three trials for which there were three or more patients expected to experience grade 4 toxicity using design 1 were T86-004, T88-114, and T91-018. These three trials are characterized by a combination of very steep dose–toxicity curves [as indicated by small values of $(K_3-K_2)/\ln(1.4)$] and relatively large amounts of intrapatient variability. With designs 2 or 4, the increase in the expected number of grade 4 toxic effects compared with design 1 is one patient or fewer in 12 of the 20 trials. The increase is greater than three patients in the three trials (T90-156, T91-018, and T92-108) characterized by very steep dose–toxicity curves. The increase in incidence of grade 4 toxicity was greater for design 3 than for designs 2 or 4.

The results presented above combined the conventional cohort escalation design 1 with the conventional intrapatient dose-modification option A. Combining design 1 with intrapatient option B has no effect on the number of patients or number of cohorts compared with 1A. It reduces the average number of patients with grade 0-1 as their highest level of toxicity from 23.3 to 19.3, but this is still not competitive with the numbers for designs 2 through 4 using option B.

Combining designs 2 through 4 with option A also has little or no effect on the number of patients or cohorts required compared with the same design using option B. In each case, about one fewer patient on average experiences grade 3 toxicity using option A than the same design with option B (5.2, 5.7, and 5.4 for 2A, 3A, and 4A, respectively). The expected number of patients with grade 4 toxicity is reduced on average by 0.4-1.1 patient (3.0, 4.3, and 3.2 for designs 2B, 3B, and 4B to 2.2, 3.2, and 2.8, respectively, for designs 2A, 3A, and 4A). The average number of patients with grade 0-1 toxicity is increased by about
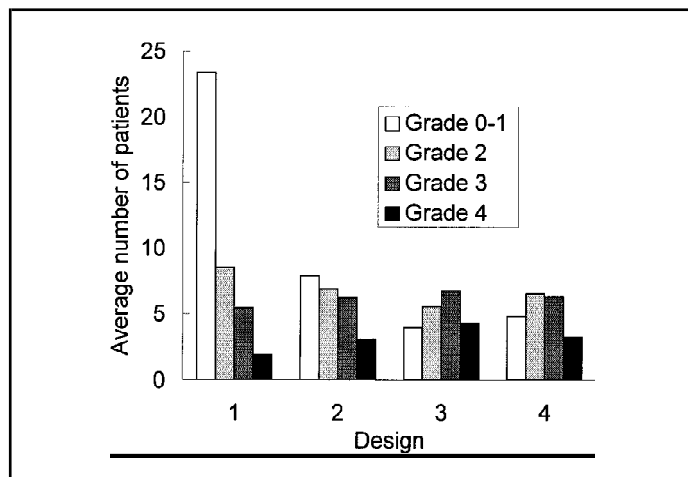
2.4-3.0 patients on average (10.3, 6.5, and 7.0 for 2A, 3A, and 4A, respectively, compared with 7.9, 3.9, and 4.8). Much of the reduction in numbers of undertreated patients is achieved with designs 2A, 3A, and 4A compared with the standard design 1A, and they result in somewhat less grade 3-4 toxicity than the designs using dose titration. They are particularly attractive when there is preclinical concern about cumulative toxicity. They do not, however, provide patients accrued early in the trial a full opportunity to be treated at a dose that provides the greatest opportunity for benefit. Also, in situations where interpatient variability is substantial relative to $(K_2 - K_1)/\ln 1.4$ and intrapatient variability is small, designs without intrapatient dose escalation will not give each patient as much opportunity to be treated at a dose level appropriate to her particular level of drug tolerance and thus will be much less effective than designs with dose titration. Such combinations of parameters are not frequent in Table 1, but smaller values of σ_ε may be more prevalent with prospective use of accelerated designs.

Example

We generated one set of data for a clinical trial with the use of design 4 and the parameter values estimated from the actual data for trial T88-127 of chloroquinoxaline sulfonamide. Table 2 shows the data generated using these parameters. The first column lists patient sequence numbers. Each row of the table corresponds to a single patient. The numbers in a row represent the grades of toxicity experienced by that patient during her three courses of therapy. The columns correspond to dose levels, and the levels are labeled at the top of the columns.

The first patient received dose level 1 in her first course, and this resulted in toxicity grade 0 or 1. The table records this as a 0 because our simulations and analysis do not distinguish between grades 0 and 1. Since design 4 is used in this example, the first patient had her dose escalated by two steps for her second course, and she again showed grade 0-1 toxicity. Consequently, she received dose level 5 for her third course. She again showed no toxicity.

Since patient 1 had no toxicity in her first course, patient 2 started at dose level 3. Our simulations assumed that the time between patient entries was the same as the length of a single treatment course. Patient 2 also did not show any toxicity in her first course, and her dose was escalated two steps to level 5 for her second course. At that same time, patient 3 started at dose level 5.

The first toxicity observed was grade 2, which occurred in the second course of therapy for patient 4 at dose level 9. Hence, her dose was not escalated for her third course.

Patient 6 had grade 2 toxicity during her first course that was at dose level 11. She was kept at dose level 11 for her second course, but it resulted in no toxicity. Consequently, her dose was escalated to level 12 for her third course. It was escalated only a single dose step because the grade 2 toxicity she experienced during her first course was the second instance of grade 2 toxicity during the trial. This ended the rapid escalation phase of the design. Consequently, the cohort started at dose level 11 was expanded to three new patients started at that dose. The single dose escalations of three patients per cohort continued until the second patient started at dose level 15, patient 19, experienced grade 3 toxicity. That cohort is therefore expanded to six patients. Patient 22 experienced grade 4 toxicity in her first course at dose level 15, and hence the escalation of starting dose for new cohorts of patients stops. Three additional patients started

Table 2. Sequence of dose escalations and toxicity grades for patients treated in simulated phase I trial using design 4*

| Patient No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
													Dose step†				
1	0		0		0												
2		0			0		0										
3			0				0		0								
4							0_1		$2_2/0_3$								
5									0		0		0				
6											$2_1/0_2$	0_3					
7											0	0	0				
8											0	0	0				
9											0_1	0_2		2_3			
10											0_1	0_2		3_3			
11											0	0		0			
12													$2_1/2_2/2_3$				
13													0	0	0		
14													$2_1/2_2/2_3$				
15														0	0	0	
16														0_1	0_2	2_3	
17														0_1	$2_2/2_3$		
18															0_1	0_2	3_3
19														0_2	$3_1/2_3$		
20															$2_1/2_2/2_3$		
21															$2_1/2_2/0_3$		
22														$2_2/2_3$	4_1		
23															$2_1/2_2/0_3$		
24														0	0	0	
25														$0_1/2_3$	4_2		
26														0_1	0_2	2_3	

*Subscript = treatment course.
†Units are just the sequentially numbered dose steps. Level 1 = starting dose. Level 2 corresponds to a dose 40% greater than the starting dose.

on the next lower dose level, level 14. No patients experienced DLT at that level, and hence accrual to the trial was completed. The traditional recommended phase II dose would be level 14.

We fit the model to the data of Table 2 obtaining the following maximum likelihood estimates with 90% confidence intervals (CIs): K_1 is estimated as 7.4 (90% CI = 6.8-7.9) instead of the true value of 7.5; $(K_2 - K_1)/.34$ is estimated as 4.1 (90% CI = 1.3-7.0) instead of the true value of 4.6; $(K_3 - K_2)/0.34$ is estimated as 1.4 (90% CI = 0-2.9) instead of 2.9; α is estimated as 0 (90% CI = 0-0.67) with the true value of 0; σ_β is estimated as 0.71 (90% CI = 0.40-1.25), with a true value of 0.62; and σ_ε is estimated as 0.83 (90% CI = 0.37-1.84), with a true value of 0.90. In this example, there is good agreement between the estimates obtained from fitting the model and the true values used to generate the data example. The CIs are based on the usual normal approximations to the maximum likelihood estimates of the K's, $\log \sigma_\beta$, and $\log \sigma_\varepsilon$ and on the approximate chi-squared distribution of the logarithm of the likelihood ratio statistic as a function of α.

There is no evidence of cumulative toxicity because the alpha parameter is estimated as zero. There appears to be a substantial amount of both interpatient variability and intrapatient variability. The standard deviations are large, relative to the logarithm of
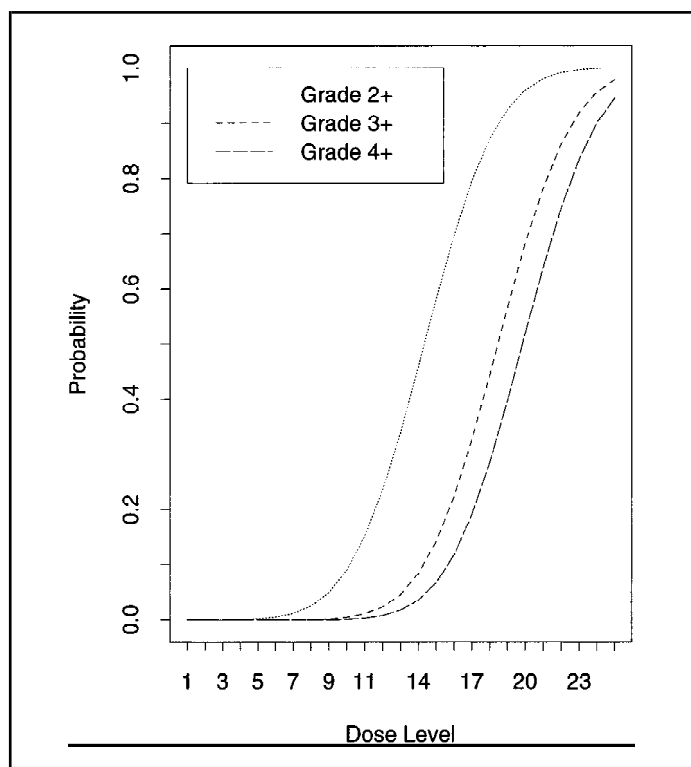
Fig. 5. Probability of toxicity of each grade level in a single course of treatment as a function of dose level. Probabilities are averaged over the population of patients. Probability curves are computed from model 1 using maximum likelihood estimates of model parameters. Specifically, the probability of grade 2+ toxicity with dose d and cumulative dose for previous courses of D is

$$\Phi \left(\frac{\log(d + \alpha D) + \mu_\beta - K_1}{\sqrt{\sigma_\beta^2 + \sigma_\varepsilon^2}} \right),$$

where Φ denotes the cumulative standard normal distribution function. For computing probability of grade 3+ or grade 4 toxicity, replace K_1 by K_2 and K_3, respectively.

the dose step $\ln(1.4)$ that is about 0.34. The K_1 and K_2 values appear to be well separated, but the K_2 and K_3 values are close.

Fig. 5 shows the probability of grade 2 or worse toxicity as a function of dose level and similar functions for the probability of grade 3 or worse toxicity and of grade 4 toxicity. These functions were computed by use of the model parameters estimated from the simulated data. From these graphs, one can estimate the dose level associated with any target level of any grade of toxicity. If one were to recommend a single dose level, the recommendation should reflect the distance between the grade 3+ curve and grade 4+ curve in Fig. 5. At dose level 17, the model estimates that 19% of the patients will experience grade 4 toxicity. At dose level 16, the probability of grade 4 toxicity is reduced to 12%, the probability of grade 3+ toxicity is 22%, and the probability of grade 2+ toxicity is 70%.

The functions in Fig. 5 do not give a clear picture of interpatient differences. Fig. 6 shows curves of the probability of grade 2+, 3+, and 4+ toxicity for three representative patients. The middle graph is for a patient whose β value equals the mean μ_β. The upper graph is for a patient whose β value is one standard deviation below the mean; i.e., $\mu_\beta - \sigma_\beta$. The lower graph is for a patient with $\beta = \mu_\beta + \sigma_\beta$. Dose levels 16 or 17 may be reasonable for the patient represented by the middle graph. For the patient represented by the upper graph, dose level 19 would be more appropriate. For the patient represented by the lower graph, dose level 14 or 15 would be more appropriate. This graph illustrates the substantial interpatient variability in the toxic response to this drug in this patient population. The separation between the grade 2+ and grade 3+ curves here and in Fig. 5 indicates the ability to effectively titrate patients to grade 2 toxicity. The closeness of the grade 3+ and grade 4+ curves indicates that doses that give grade 3 toxicity overlap substantially with those that give grade 4 toxicity. Use of any fixed dose for all patients is problematic, since any dose both overtreats and undertreats some patients. This is the principal conclusion of the data analysis.

Discussion

The new designs described here appear to accomplish several objectives. They reduce the number of patients potentially undertreated. Some of these designs also reduce the duration of trials by doubling the dose until toxicity develops. These approaches also improve the information yield of phase I trials. They provide for estimation of the population distribution of the MTD and may also provide a statistical estimate of the degree of cumulative toxicity.

We have addressed phase I trials in which patients may receive more than one course of treatment. Not all phase I trials are of this type. Even in trials of this type, many patients remain in the study for only one or two courses of treatment because of tumor progression. This limits the information available for analysis. Patients may be able to remain in the study longer with accelerated titration designs because use of intrapatient dose escalation provides greater opportunity for therapeutic benefit. The reduced risk of design 4 compared with design 3 was based on using information from the second and third courses in determining when to stop the initial accelerated stage. This additional protection can be assured with fewer courses of treatment per patient by requiring that when the first instance of grade 2

Fig. 6. Probability toxicity of each grade level in a single course of treatment as a function of dose level for individual patients. Probabilities are not averaged over the population of patients but are computed separately for the average patient (middle panel with $\beta_i = \mu_\beta$), the patient with reduced sensitivity to the toxic effects of the compound (upper panel with $\beta_i = \mu_\beta - \sigma_\beta$), and the patient with increased sensitivity to the toxic effects of the compound (lower panel with $\beta_i = \mu_\beta + \sigma_\beta$). Probability curves are computed from model 1 using maximum likelihood estimates of model parameters. Specifically, the probability of grade 2+ toxicity with dose d and cumulative dose for previous courses of D for a patient with value β_i is

$$\Phi\left(\frac{log(d + \alpha D) + \beta_i - K_1}{\sigma_\varepsilon}\right),$$

where Φ denotes the cumulative standard normal distribution function. For computing probability of grade 3+ or grade 4 toxicity, replace K_1 by K_2 and K_3, respectively.
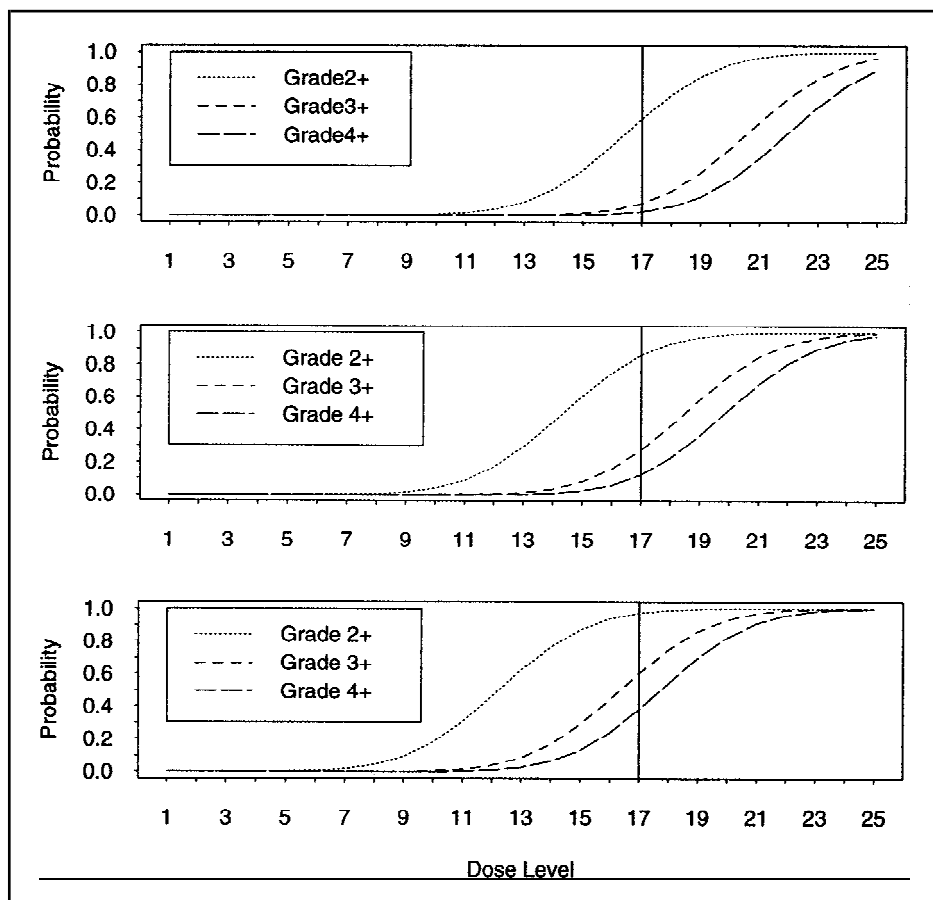
toxicity occurs, two other patients be treated at that same dose without grade 2 toxicity before the dose is doubled. This may be satisfied by later courses at escalated doses in previous patients or may require starting a new patient at the same dose as the one who experienced grade 2 toxicity. This modification is not needed for design 2, since it uses only first-course toxicity for determining when to terminate the accelerated stage and uses smaller dose steps. For designs without intrapatient dose escalation, this modification would increase the number of patients treated at lower doses and may extend the time to completion.

In these simulations, we used the conventional stopping rule with all designs for consistency. The study stopped when two patients experienced DLT at a dose level, and six patients were treated at the next lower dose level with no more than one patient experiencing DLT. For the new designs, the population distribution of MTDs is estimated, and there is nothing special about the highest dose at which fewer than two patients experienced DLT. One might, therefore, continue entering patients beyond the usual stopping point to refine the estimates of the population distributions. In fact, the entire second stage of sampling could use a model-based or Bayesian approach to selecting the first-course dose for each patient. Simple up-down phase I designs with cohort sizes other than three to six patients are also sometimes used when the amount of DLT that is to be tolerated is much less than 33%.

We have analyzed the results of 20 actual phase I trials by use of the model described above for the generation of simulated trial data. Other models could be used in place of the expression shown in equation 1 and, in particular, other approaches to the

modeling of cumulative toxicity may be more appropriate in specific trials.

Use of an accelerated titration design requires careful definition of the level of toxicity considered dose limiting and the level considered sufficiently low (e.g., none-mild) that intrapatient dose escalation is acceptable. These definitions must be made for each organ system. In the simulations, we tabulated the incidence of unacceptable or grade 4 toxicity, but this is not necessary in using an accelerated titration design. The dose escalation and de-escalation decisions that must be made during the trial depend on distinguishing none-mild toxicity from moderate toxicity and on distinguishing moderate toxicity from DLT. These definitions may be protocol specific. The tracking of toxicity over multiple treatment courses and the use of intrapatient titrations require careful patient management. However, the result will enhance the likelihood that patients receive therapeutic dosing and increase the useful information obtained from each treated patient.

The approach to design and analysis of phase I trials described in this article will help identify when there is large interpatient variability in sensitivity to the toxic effects of a drug. If interpatient variability is small, a fixed-dose regimen can be used in phase II trials, and few patients will be either overdosed or underdosed. Mick et al. (18) have described important sources of interpatient and intrapatient variability that might be usefully incorporated into the model. Further improvement might result from modeling toxicity separately by organ system.

Pharmacokinetic differences are sometimes an important source of interpatient variability. In such cases, it may be advis-

able to attempt to control systemic exposure rather than dose. If drug clearance can be predicted by use of baseline patient characteristics such as liver or renal function, then the dose needed to achieve the targeted concentration can be estimated. Otherwise, an adaptive dosing scheme may be needed to achieve a target exposure. If the drug is delivered by a prolonged infusion, one may adapt the infusion rate based on estimates of pharmacokinetic parameters to target systemic exposure levels. The accelerated titration designs described here then may be applied with the only change being the use of exposure levels rather than dose levels. When prolonged infusions are not used, it may not be feasible to deliver a target exposure level during the same course of treatment in which pharmacokinetic parameters are estimated. It still may be possible, however, to use parameters estimated in the first course of treatment for the titration of exposure in subsequent courses.

Accelerated titration designs are more aggressive than standard approaches and, therefore, may be associated with more risk. The simulations were performed with a very wide range of model parameters and suggest that the risks appear acceptable for designs 2 and 4. We believe that these designs are appropriate for clinical testing. For drugs that exhibit preclinical evidence of cumulative toxicity, special caution in the conduct of any type of phase I trial is needed. Accelerated designs without intrapatient dose escalations achieve most of the advantages of accelerated titration designs, with little or no increase in risk compared with the standard design 1A. However, they do not provide as great a reduction in the number of undertreated patients and, in particular, do not provide patients accrued early in the trial or those who have an especially high individual tolerance for the drug as much opportunity as do titration designs to be treated at a dose that provides the greatest opportunity for benefit. We hope to sponsor phase I clinical trials to provide prospective evaluation of these new approaches.

Appendix

Four designs were evaluated as follows. **Design 1:** cohorts of three new patients per dose level. If one of three patients experiences DLT in the first course, expand the cohort to six patients. Intrapatient escalation option A. **Design 2:** cohorts of one new patient per dose level. When the first instance of first course DLT is observed, or the second instance of first course grade 2 toxicity of any type, expand the cohort for current dose level and revert to use of design 1 for all further cohorts. Intrapatient escalation option B. **Design 3:** same as design 2, except that double dose steps are used during the initial accelerated stage of the trial (both for between-patient and within-patient escalations). Intrapatient escalation option B. **Design 4:** cohorts of one new patient per dose level and double dose steps are used during the initial accelerated stage of the trial. When the first instance of DLT is observed at any course or the second instance of any course grade 2 toxicity of any type, expand the cohort for current dose level and revert to use of design 1 for all further cohorts. Intrapatient escalation option B.

The intrapatient dose modification options are defined as follows:

Option A: no within-patient dose escalation. De-escalate if grade 3 or worse toxicity at previous course.

Option B: Escalate if grade 0-1 toxicity at previous course. De-escalate if grade 3 or worse toxicity at previous course.

References

(*1*) Ratain MJ, Mick R, Schilsky RL, Siegler M. Statistical and ethical issues in the design and conduct of phase I and II clinical trials of new anticancer agents. J Natl Cancer Inst 1993;85:1637-43.

(*2*) Estey E, Hoth D, Simon R, Marsoni S, Leyland-Jones B, Wittes R. Therapeutic response in phase I trials of antineoplastic agents. Cancer Treat Rep 1986;70:1105-15.

(*3*) Collins JM, Zaharko DS, Dedrick RL, Chabner BA. Potential roles for preclinical pharmacology in phase I clinical trials. Cancer Treat Rep 1986; 70:73-80.

(*4*) Storer BE. Design and analysis of phase I clinical trials. Biometrics 1989; 45:925-37.

(*5*) Eichhorn BH, Zacks S. Sequential search for an optimal dosage. J Am Stat Assoc 1973;68:594-8.

(*6*) O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. Biometrics 1990;46:33-48.

(*7*) Faries D. The modified continual reassessment method for phase I cancer clinical trials. American Statistical Association 1991 Proceedings of the Biopharmaceutical Section. Alexandria (VA): American Statistical Association, 1991:269-73.

(*8*) Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. Stat Med 1995;14: 1149-61.

(*9*) Moller S. An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses. Stat Med 1995;14: 911-22.

(*10*) Piantadosi S, Liu G. Improved designs for dose escalation studies using pharmacokinetic measurements. Stat Med 1996;15:1605-18.

(*11*) Mick R, Ratain MJ. Model-guided determination of maximum tolerated dose in phase I clinical trials: evidence for increased precision. J Natl Cancer Inst 1993;85:217-23.

(*12*) Sheiner LB, Beal SL, Sambol NC. Study designs for dose-ranging. Clin Pharmacol Ther 1989;46:63-77.

(*13*) Sheiner LB, Hashimoto Y, Beal SL. A simulation study comparing designs for dose ranging. Stat Med 1991;10:303-21.

(*14*) Chuang C. The analysis of a titration study. Stat Med 1987;6:583-90.

(*15*) Shih WJ, Gould AL, Hwang IK. The analysis of titration studies in phase III clinical trials. Stat Med 1989;8:583-91.

(*16*) Chou TC, 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. Eur J Biochem 1981;115: 207-16.

(*17*) Benet L, Oie S, Schwartz J. Design and optimization of dosage regimens; pharmacokinetic data. In: Hardman J, Limbird L, Molinoff P, et al., editors. The pharmacological basis of therapeutics. New York: McGraw-Hill, 1996: 1707-92.

(*18*) Mick R, Lane N, Daugherty C, Ratain MJ. Physician-determined patient risk of toxic effects: impact on enrollment and decision making in phase I cancer trials. J Natl Cancer Inst 1994;86:1685-93.