Research paper

Flexible, rule-based dose escalation: The cohort-sequence design

Shuang Li [a], Xian-Jin Xie [b], Daniel F. Heitjan [a,c,*]

[a] *Department of Statistical Science, Southern Methodist University, Dallas, TX, USA*
[b] *College of Dentistry and College of Public Health, University of Iowa, Iowa City, IA, USA*
[c] *Department of Population & Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA*

ARTICLE INFO

ABSTRACT

Phase I oncology trials seek to acquire preliminary information on the safety of novel treatments. In current practice, most such trials employ rule-based designs that determine whether to escalate the dose using data from the current dose only. The most popular of these, the $3 + 3$, is simple and familiar but inflexible and inefficient. We propose a rule-based design that addresses these deficiencies. Our method, which we denote the *cohort-sequence* design, is defined by a sequence of J increasing cohort sizes $n = (n_1, ..., n_J)$ and corresponding critical values $b = (b_1, ..., b_J)$. The idea is to begin with a small cohort size n_1 and escalate through the planned doses, increasing the cohort size when we encounter toxicities. By selection of J and a safety threshold tuning parameter θ, one can create a design that will efficiently identify a target toxicity rate, potentially including a built-in dose-expansion cohort. We compared our designs to the $3 + 3$ under a range of toxicity scenarios, observing that our approach generally rapidly identifies an MTD without enrolling patients unnecessarily at low doses where both toxicity and response rates are likely to be low. We have implemented the design in the R package *cohortsequence*.

1. Introduction

We acquire preliminary information on the safety of a novel treatment through the conduct of a *phase I trial*. Statisticians commonly formulate the objective of such a trial as the identification of the treatment's maximum tolerated dose (MTD) — that is, the dose that gives the highest acceptable rate of dose-limiting toxicities (DLTs). The typical design calls for enrolling subjects in dose cohorts, starting from a low dose that is believed to be safe and increasing after each cohort until encountering a designated level of toxicity.

Most phase I trials employ *rule-based* designs that use data from only the current cohort to decide what the next step will be. The most popular such design is the $3 + 3$, a variant of the up-and-down rule [1–6]. Some alternatives to the $3 + 3$ include the $A + B$ design [4], which generalizes the cohort sizes in the $3 + 3$; the accelerated titration design, which starts with one-patient cohorts and reverts to the $3 + 3$ plan once toxicities appear [7]; and designs based on intervals around the target toxicity probability [8,9] such as the toxicity probability interval (TPI), modified TPI (mTPI), and Bayesian optimal interval design (BOIN), which use Bayesian criteria to select the next dose [3,5,10,11]. Yet despite the availability of these and other alternative designs, many with excellent statistical properties, as recently as 2009 nearly 97% of phase I

trials used the $3 + 3$ [1]. More recent publications also note the limited use of innovative designs in clinical trial practice [12,13].

The $3 + 3$ design is simple and familiar but has notable defects: First, for commonly encountered toxicity profiles, the $3 + 3$ most often selects doses with DLT rates in the range 20%–25%, well below the typical nominal target of 30%–35% [14]. Second, most patients in a $3 + 3$ trial receive doses that give low rates of both toxicity and therapeutic effect; a more efficient design would escalate quickly past these to reach doses that are of greater interest. Third, the maximum number of patients that a $3 + 3$ enrolls at the final dose is 6, implying that it can obtain only limited information about toxicity and efficacy at the purported MTD. Thus it has become common to augment the phase I trial with a *dose-expansion cohort* — an additional group of patients who receive treatment at the identified MTD. Often, the choice of sample size for this additional cohort is arbitrary [15], and in any event an expansion cohort is of no value if the trial mis-estimates the MTD.

Our objective in this article is to describe a novel rule-based dose-finding design that avoids the problems of the $3 + 3$ while preserving, to the extent possible, its simplicity. Our method, which we denote the *cohort-sequence* design, permits tuning of the cohort sizes and critical values to reflect the targeted DLT rate. It improves efficiency by focusing enrollment at doses where toxicities are likely to occur, thereby creating

---

larger cohorts in the vicinity of the MTD and obviating the need to append an arbitrarily sized dose-expansion cohort.

2. Methods

2.1. Escalation plan for the cohort-sequence design

The cohort-sequence design consists of a sequence of cohort sizes $n = (n_1, ..., n_J)$ and corresponding DLT critical values $b = (b_1, ..., b_J)$ that signal whether to escalate, de-escalate, or enroll more subjects at the current dose. The notion is to begin with a small cohort size n_1 and escalate through the planned series of increasing doses $D = (D_1, ..., D_m)$, raising the cohort size as we begin to encounter toxicities. Specifically, when enrolling subjects at dose D_i with cohort size n_j, the decision to escalate, add more at the current dose, or de-escalate hinges on whether the observed DLT count in the cohort falls below, equals, or exceeds the corresponding critical value b_j. If the DLT count at dose D_i exceeds b_j, we stop treating the cohort at that dose to avoid excessive toxicities. A possible value for the sequence of cohort sizes would be $n = (1, 3, 5, 8, 10)$, with corresponding sequence of critical values $b = (1, 2, 3, 4, 5)$.

Fig. 1 displays a flow chart for our design. Suppose that the current dose level is D_i and that our current cohort size is n_j with corresponding critical value b_j. We enroll up to n_j subjects at this dose and observe the number of DLTs as X_i. If $X_i < b_j$, we deem the current dose to be safe, and we escalate and enroll the next cohort at dose D_{i+1} with the same cohort size n_j and critical value b_j. If $X_i > b_j$, we deem the current dose unsafe, and we enroll the next cohort at the next lower dose D_{i-1} with the terminal cohort size n_J and corresponding critical value b_J. If $X_i = b_j$, we deem the current dose as potentially, but not certainly, toxic, and we increase the cohort size to n_{j+1} and the corresponding critical value to b_{j+1}, enrolling additional subjects at this dose until the cohort is filled and we can again evaluate the safety data. The escalation/de-escalation continues in this way until a safe dose is achieved ($X_i < b_J$) after a de-escalation, or the trial de-escalates to the lowest dose. If we escalate to the highest dose D_m with cohort size n_j, and observe $X_m < b_j$, then we increase the cohort size to n_J and the critical value to b_J. We estimate the MTD as the highest safe dose evaluated. Alternatively, one can specify a total number of patients and stop when all have received treatment, again identifying the highest safe dose as the MTD.

2.2. Identification of the critical value for a designated cohort size

We select the pairs $(n_j, b_j), j = 1, ..., J$ to represent a comparable level of certainty about the DLT rate at the given dose. That is, for each n_j, the selected b_j is one that identifies the current dose as safe, unsafe, or indeterminate, by a criterion that is common across cohort sizes. To accomplish this, we first select a safety threshold, denoted $\theta \in (0, 1)$. The principle is that if we are reasonably sure that the DLT rate at dose D_i is below θ, we escalate; if we are reasonably sure that the DLT rate at D_i exceeds θ, we de-escalate; otherwise, we collect more data at D_i. Having chosen θ, we can either compute the critical value b_j from a selected cohort size n_j, or *vice versa*.

We first demonstrate the computation of b_j from n_j. We evaluate uncertainty about the toxicity level using Bayesian posterior probabilities. At dose level D_i, the number of DLTs X_i is binomial with parameters (N_i, τ_i), where N_i and $\tau_i = \tau(D_i)$ represent the cohort size and toxicity probability, respectively. We assume a Beta(1,4) prior for τ_i, which updates to Beta$(1 + X_i, 4 + N_i - X_i)$ after X_i DLTs in N_i trials. We use this prior because it represents the situation where there are $X = 0$ events in a previous $N = 3$ patients starting from a uniform prior, which approximates the level of uncertainty that one would express before examining an untested dose level. As in the TPI and mTPI designs [3,5, 10], we use the same prior for all dose levels because we have essentially no prior data at any dose level.

Starting from a fixed n_j, to determine the critical level b_j we compute the posterior probability that τ_i exceeds the safety threshold θ, which we denote

$$f(X_i, N_i, \theta) \equiv \Pr[\tau_i > \theta | X_i, N_i].$$

We fix a threshold of 10% for this posterior probability. For example, if $f(X_i, N_i, \theta)$ is below 10%, then we deem the dose level safe for escalation. If $f(X_i, N_i, \theta)$ is well above 10%, we de-escalate. If $f(X_i, N_i, \theta)$ is above 10% but $f(X_i - 1, N_i, \theta)$ is below 10%, we collect more data at the current dose. We select b_j by identifying the value of X_i such that $f(X_i, N_i, \theta) > 10\%$ and $f(X_i - 1, N_i, \theta) \le 10\%$.

2.3. Identification of the cohort size for a designated critical value

Alternatively, one can select a sequence of b_j values and then calculate the corresponding cohort sizes n_j. To avoid ambiguities, the b_j
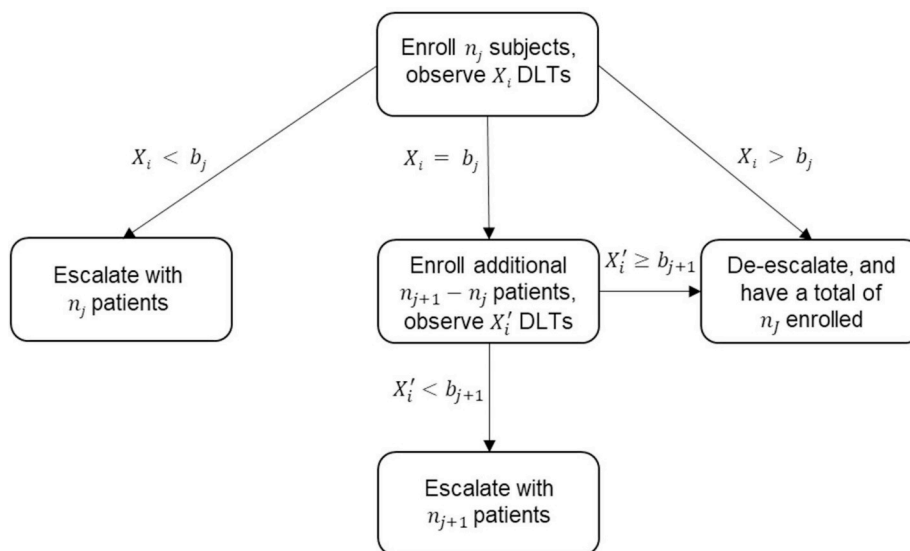

Fig. 1. Flow chart for the cohort-sequence design.

values should constitute an increasing sequence; $b = (1, 2, ..., J)$ is a natural choice. We start by computing $f(b_j - 1, x, \theta)$ for $x \geq b_j$, which decreases as x increases. Then n_j is the smallest x such that $f(b_j - 1, x, \theta) \leq 10\%$, and $f(b_j, x, \theta) > 10\%$. Fig. 2 illustrates the process of identifying the cohort size.

2.4. Selection of the number of cohort sizes

The final element of the design is the number of cohort sizes, which we denote J. As a rule, for a fixed b_j, the smaller the value of θ, the larger will be the corresponding n_j. Therefore, for small values of θ it is preferable to choose a smaller J, lest the total sample size be excessive. In the next subsection we demonstrate some feasible choices of θ and J.

2.5. Example designs

Flexibility of the design derives from the safety threshold θ. We note that one should think of θ as a tuning parameter and not a target DLT rate; in simulations (see Section 3 below), we show that θ typically exceeds the modal DLT rate by 10%–20%. For example, $\theta = 0.25$ leads to a low maximum DLT rate; $\theta = 0.35$ to a moderate rate; and $\theta = 0.50$ to a high rate. Thus a strategy for identifying a design is to simulate frequency properties under a likely dose-response curve for a range of values of θ and J, selecting the pair that gives the desired target DLT rate with a feasible sample size.

We use the notation $CS(100\theta; n_1, ..., n_J)$ to denote a cohort-sequence design using θ as the toxicity threshold and $n_1, ..., n_J$ as the sequence of cohort sizes, assuming by default that $b = (1, 2, ..., J)$. For cohort-sequence designs with high maximum acceptable DLT rate we set $J = 5$ and $\theta = 0.5$, which leads to cohort sizes $n = (1, 3, 5, 8, 10)$. We denote this design CS(50; 1,3,5,8,10), or henceforth CS(50). For moderate maximum acceptable DLT rate, we set $J = 2$ and $\theta = 0.35$ and derive $n = (2, 6)$, designated CS(35; 2,6). For low maximum acceptable DLT rate, we set $\theta = 0.25$ and $J = 2$ and compute $n = (5, 11)$, designated CS (25; 5,11). Alternatively, letting $\theta = 0.4$ and $b = (1, 2, 3)$, the cohort sizes are $n = (3, 6, 9)$, which is similar to the 3 + 3 with an expansion cohort of size 3; we designate this design CS(40; 3,6,9). We present these n, b combinations in Table 1.

With $J > 5$ or $\theta < 0.3$, the possible total number of subjects in the study can exceed the sample sizes that are typical for phase I trials. On the other hand, with $J < 5$, one may not achieve the numbers of subjects at the candidate MTD that we typically observe in practice when θ is large. With $J = 5$, the cohort size at the identified MTD is similar to the sizes of typical dose-expansion cohorts, making it unnecessary to enroll additional patients at those doses.

We have implemented the cohort-sequence design in the R package *cohortsequence*, which computes $f(X_i, N_i, \theta)$ and can calculate b_j from n_j and *vice versa*. The package also provides a function to simulate the performance of a cohort-sequence design for a designated dose-response scenario.
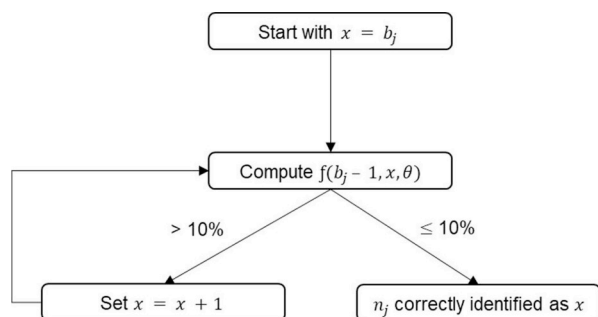

Fig. 2. Flow chart for determining the cohort size based on specified DLTs (X), under safety threshold θ.

Table 1
Decision rules for cohort-sequence designs used in the simulation.

DLT limit (b_j)	Sequence of cohort sizes (n)			
	CS(25; 5,11)	CS(35; 2,6)	CS(40; 3,6,9)	CS(50)
1	5	2	3	1
2	11	6	6	3
3			9	5
4				8
5				10

3. Simulation

3.1. Design

Ahn conducted simulations under four dose/toxicity scenarios to compare the 3 + 3 with variants of the model-based continual reassessment method (CRM) [16]. We have used the same scenarios, plus two others, to evaluate the frequency properties of cohort-sequence designs with $\theta \in \{0.25, 0.35, 0.4, 0.5\}$. In our additional scenarios, the toxicity rises abruptly to a high level, a situation not considered by Ahn. Our simulations terminate cohorts as soon as the number of DLTs is high enough to signal toxicity, a practice that improves efficiency and eliminates inadmissible toxicity.

Our simulations also include the mTPI, BOIN, and CRM designs [3,5, 10,11,17]. The mTPI is guided by the posterior probability of a pre-specified toxicity probability interval around the target toxicity, whereas the BOIN compares observed fractions of toxicities with boundaries derived from a pre-specified toxicity tolerance interval. The CRM design identifies the MTD by iteratively updating estimates of the dose-toxicity model and selecting the dose that gives the targeted toxicity rate.

We evaluated the performance of eleven dose-escalation designs:
1. The traditional 3 + 3.
2. The 3 + 3 with an expansion cohort of size 3, with the stopping bound for safety in the expansion cohort set to 2 [15]; we denote this design as 3 + 3@9,2. With this design, we de-escalate from an unsafe expansion cohort and enroll another until the final dose is safe.
3. CS(40; 3,6,9): A cohort-sequence design that is similar to the 3 + 3@9,2.
4. CS(25; 5,11): Suitable for a low target DLT rate.
5. CS(35; 2,6): Suitable for a moderate target DLT rate.
6. CS(50; 1,3,5): Suitable for a higher target DLT rate with fewer patients.
7. CS(50): Suitable for a higher target DLT and including a built-in expansion cohort.
8. mTPI: The mTPI design with target toxicity of 30%, cohort size of 3, maximum sample size being the rounded result from the 3 + 3 design for each corresponding scenario, and other parameters set at defaults [3].
9. mTPI2: Same as mTPI, but with target toxicity 35%.
10. BOIN: The BOIN design with target toxicity 30%, cohort size 3, sample size same as mTPI, and other parameters at defaults [11].
11. CRM: The CRM design with target toxicity 30%, cohort size 1, same sample size as mTPI. The prior distribution for DLT probabilities is $(0.01, 0.1, 0.2, ..., 0.8)$ for Scenarios 1–4, and $(0.01, 0.4, 0.8)$ for Scenarios 5 and 6.

We repeated the simulation 5000 times. When the lowest dose was rejected as toxic, the estimated MTD was designated as dose level 0. In this case, the number of patients treated at the estimated MTD is 0, and we do not enroll an expansion cohort. We compared performance of the designs using four criteria:

1. The proportion of times each dose was recommended as the MTD;
2. the fraction of patients treated at each dose;
3. the average number of patients enrolled; and
4. the average proportion of patients experiencing a DLT.

One can specify the MTD based on a target toxicity, then compare the ability of the designs to identify it; that is, any scenario can be re-used to reflect different target toxicities. An ideal design should have accurate estimation of the MTD (Criterion 1); better patient allocation (Criterion 2); a small number of patients (Criterion 3); and a relatively large overall toxicity fraction (Criterion 4). We emphasize that Criterion 4 reflects efficiency, in the sense that a design that skips past safe (and ineffective) doses will give a higher overall toxicity rate than a design that lingers at low doses. On the other hand, a design that gives a high overall toxicity rate while treating many subjects at doses in excess of the MTD (i.e., not satisfying Criteria 1 and 2) is not desirable.

Because BOIN and mTPI designs give similar results, henceforth we do not present BOIN data.

3.2. Results

Tables 2–5 display simulated frequency properties of the designs applied to Ahn's scenarios, and Tables 6 and 7 display results from the additional scenarios. When the target toxicity is $\leq 10\%$, the CS(25; 5,11) design gives the correct estimate most frequently. When the target toxicity is between 10% and 25%, the prediction accuracy for CS(35) is higher. When the target toxicity exceeds 25%, designs with $\theta = 0.5$ lead to correct estimates most often. Cohort-sequence designs generally treat lower fractions of subjects at low, safe doses. An exception is the CS(40; 3,6,9), which closely mimics the behavior of $3 + 3$.

The $3 + 3@9,2$ and CS(40; 3,6,9) designs require 4 to 6 more patients than the traditional $3 + 3$, with the extra patients constituting a built-in dose-expansion cohort. For cohort-sequence designs with $\theta = 0.5$, the average number of patients with $J = 3$ is smaller than with $J = 5$, although other frequency properties are similar. The CS(50; 1,3,5) in particular requires fewer patients than the $3 + 3$. CS(35; 2,6) performs similarly to $3 + 3$ in terms of MTD recommendation and patient allocation, but it requires 3 fewer patients in toxicity Scenario 4 and 1 fewer patient in the other scenarios.

With a higher value of the tuning parameter θ, the realized toxicity fraction is typically higher. Yet even for the CS(50) designs, which deliberately target higher toxicity rates, the fractions of subjects experiencing toxicity are less than 35% under all scenarios except for Scenario 5, where the lowest dose is too toxic. The toxicity percentages for the traditional $3 + 3$ design are similar to those for CS(35; 2,6) and CS(40; 3,6,9).

The CS(50) design allows for larger cohort sizes than the $3 + 3$, but the realized sample size is typically smaller than that of the $3 + 3$ with an expansion cohort of 3. Moreover, the fraction of toxicities is greater, and it assigns more patients at higher dose levels. Nevertheless, CS(50) effectively avoids extremely toxic doses, as subjects rarely reach doses with DLT rates in excess of 50%. The CS(25; 5,11) design enrolls subjects at these highly toxic doses only in Scenarios 3 and 4, where there is a jump from 25% DLTs to 80% DLTs in one dose elevation. The CS(50) designs enroll more patients at DLT rates above 50% in Scenario 5, where the lowest dose has toxicity probability of 45%

When the target toxicity is 30%, the mTPI design outperforms CS(50) in Scenarios 1, 2, and 6, but requires larger sample size than CS(50; 1,3,5) in the first two scenarios. We can reduce the total sample size of mTPI and BOIN to smaller numbers that are multiples of 3. However, in situations like Scenario 4 with low toxicities, having $n < 15$ prevents dose-escalation from ever reaching the target MTD. With a higher target toxicity of 35%, the mTPI design is less efficient and assigns fewer patients at the MTD than CS(50) designs. The mean number of patients under mTPI lies between those of CS(50; 1,3,5) and CS(50). Compared with the mTPI designs using a fixed cohort size, CS typically gives smaller cohort sizes at low-toxicity doses. In Scenarios 2–4, CS(50) assigns a higher fraction of patients at the MTD. Therefore, under a target toxicity of 30% and above, one can use CS(50) designs for more efficient dose assignments and better predictions. This applies except in situations like Scenario 5, where the lowest dose is already excessively toxic. In that situation, we need a conservative dose-finding design with lower θ.

The CRM design performs similarly to mTPI in all the scenarios. The CS(50) designs are slightly better than the CRM in Scenarios 3 and 4. In Scenario 6, CS(50) designs are more aggressive. The total proportion of MTD estimation at dose 2 and 3 together are slightly higher than the result from the CRM design, but the overall toxicity is closer to 30%. The CRM design generally works well at estimating the target MTD. Yet our proposed design has similar performance (in Scenarios 1, 2, and 6) or even better (in Scenarios 3 and 4) in the simulation settings that we have examined.

We note that the mTPI and CRM designs that we examined were tailored to identify doses with toxicity rates of 30%, and for that they are

Table 2
Comparison of phase I designs under dose-toxicity Scenario 1.

Dose	DLT rate	Designs									
		$3 + 3$	$3 + 3@9,2$	CS(40; 3,6,9)	CS(25; 5,11)	CS(35; 2,6)	CS(50; 1,3,5)	CS(50)	mTPI	mTPI2	CRM
Proportion of recommended dose (MTD) dose levels (%)											
0		3	4	3	9	2	0	0	0	0	0
1	0.05	10	17	10	26	9	2	2	1	1	1
2	0.10	38	50	38	51	36	12	12	21	16	14
3	0.25	31	22	32	13	33	19	23	45	44	40
4	0.35	16	6	15	1	17	32	39	29	32	31
5	0.50	2	0	2	0	3	28	22	3	6	12
6	0.70	0	0	0	0	0	6	2	0	0	2
7	0.80	0	0	0	0	0	1	0	0	0	0
Proportion of patients treated (%)											
1	0.05	27	26	24	41	21	10	7	25	25	10
2	0.10	31	34	32	38	29	15	14	31	31	18
3	0.25	25	25	27	18	28	20	20	31	31	25
4	0.35	13	11	14	3	16	23	27	11	11	23
5	0.50	4	3	4	0	5	20	22	1	1	16
6	0.70	0	0	0	0	1	9	8	0	0	3
7	0.80	0	0	0	0	0	2	1	0	0	4
Average number of patients		15.3	19.9	19.6	21.3	14.5	12.8	20.0	15.0	15.0	15.0
Percent toxicity (%)		20.3	19.1	18.6	13.3	20.7	34.7	34.3	17.1	17.1	30.5

Table 3
Comparison of phase I designs under dose-toxicity Scenario 2.

Dose	DLT rate	Designs									
		3 + 3	3 + 3@9,2	CS(40; 3,6,9)	CS(25; 5,11)	CS(35; 2,6)	CS(50; 1,3,5)	CS(50)	mTPI	mTPI2	CRM
Proportion of recommended dose (MTD) dose levels (%)											
0		3	3	3	9	2	0	0	0	0	0
1	0.05	10	13	9	24	8	2	2	1	0	0
2	0.10	18	22	16	31	14	4	4	7	5	3
3	0.15	21	24	19	23	22	6	6	20	21	9
4	0.20	20	21	21	10	22	9	10	29	28	19
5	0.25	18	13	20	3	19	17	21	28	28	32
6	0.35	9	4	11	0	11	31	36	14	17	29
7	0.50	2	0	1	0	2	26	20	1	2	8
8	0.75	0	0	0	0	0	4	1	0	0	0
9	0.90	0	0	0	0	0	0	0	0	0	0
Proportion of patients treated (%)											
1	0.05	24	23	21	38	18	9	7	18	17	7
2	0.10	23	23	22	31	21	10	8	23	23	7
3	0.15	20	21	20	20	21	12	10	25	25	10
4	0.20	15	16	17	8	18	13	12	20	20	17
5	0.25	10	10	12	3	13	15	17	11	11	25
6	0.35	5	5	6	0	7	17	22	4	4	20
7	0.50	2	1	2	0	2	15	17	0	0	12
8	0.75	0	0	0	0	0	6	5	0	0	3
9	0.90	0	0	0	0	0	1	1	0	0	0
Average number of patients		18.7	23.3	24.9	24.6	17.7	15.6	22.6	21.0	21.0	21.0
Percent toxicity (%)		18.1	17.5	17.2	12.9	18.4	30.5	31.3	14.8	14.8	27.5

Table 4
Comparison of phase I designs under dose-toxicity Scenario 3.

Dose	DLT rate	Designs									
		3 + 3	3 + 3@9,2	CS(40; 3,6,9)	CS(25; 5,11)	CS(35; 2,6)	CS(50; 1,3,5)	CS(50)	mTPI	mTPI2	CRM
Proportion of recommended dose (MTD) levels (%)											
0		10	11	9	26	7	0	2	0	0	0
1	0.10	9	9	8	17	7	2	2	6	3	2
2	0.10	8	9	7	15	7	2	2	7	10	5
3	0.10	7	12	6	12	8	1	1	14	15	6
4	0.10	30	39	31	23	32	13	13	26	20	21
5	0.25	36	20	39	6	38	76	80	45	46	58
6	0.80	0	0	0	0	0	4	0	1	6	8
7	0.90	0	0	0	0	0	0	0	0	0	0
Proportion of patients treated (%)											
1	0.10	29	28	26	47	23	11	9	27	27	11
2	0.10	19	18	17	22	18	10	8	23	24	7
3	0.10	16	15	15	14	16	11	8	20	20	8
4	0.10	16	20	19	11	19	15	14	16	16	22
5	0.25	15	16	19	5	19	33	46	11	11	38
6	0.80	4	4	4	0	5	18	14	2	2	10
7	0.90	0	0	0	0	0	2	2	0	0	4
Average number of patients		18.2	22	23.5	24.0	17.2	13.3	19.0	18.0	18.0	18.0
Percent toxicity (%)		20.0	19.4	18.7	17.5	19.2	30.1	29.1	13.2	13.2	26.2

very effective. The 3 + 3 design is not tailored for any particular target, but most often finds doses that give toxicities in the range 10%–30%. Our point is that appropriately chosen CS designs out-perform 3 + 3 essentially always, while being competitive with mTPI and CRM when seeking the same MTD.

4. Application

It is generally impossible to compare designs on a "live" data set, because any real data would have arisen under a design that dictated a sequence of dose assignments that another design would not replicate. To attempt a realistic comparison of designs, we generated DLT responses using a probit model estimated from the data of Simon et al. [7] The model assumes $Y_i = \log(d_i) + \varepsilon_i$ with $\varepsilon_i \sim \mathbb{N}(0, \sigma^2)$, and registers a DLT if $Y_i > K$. We estimated the parameters to be $\sigma = 1.092$ and $K = 8.78$, which gave DLT probabilities of 1%, 5%, 14%, 32%, and 56% at doses 11, 13, 15, 17, and 19, respectively. For each subject, we generated a normal error ε_i and created a corresponding latent outcome under each dose d_i. In this way we created an ensemble of correlated data sets, one for each design.

Results appear in Table 8. All the methods estimate the MTD as either dose 15 (DLT rate 14%) or dose 17 (DLT rate 32%). The 3 + 3 with no expansion cohort yields a moderate sample size, a small fraction of DLTs, and the lowest proportion of subjects at or near the ultimate MTD. The 3 + 3@9,2 and the similar CS(40; 3,6,9) give identical results; compared to the 3 + 3, they have larger sample sizes, similar fractions of DLTs, and larger proportions of subjects treated at or near the MTD. CS(35; 2,6) and CS(50) give equal or larger proportions of DLTs than the 3+3-type

Table 5
Comparison of phase I designs under dose-toxicity Scenario 4.

Dose	DLT rate	Designs									
		3 + 3	3 + 3@9,2	CS(40; 3,6,9)	CS(25; 5,11)	CS(35; 2,6)	CS(50; 1,3,5)	CS(50)	mTPI	mTPI2	CRM
Proportion of recommended dose (MTD) dose levels (%)											
0		0	0	0	0	0	0	0	0	0	0
1	0.01	0	0	0	0	0	0	0	0	0	0
2	0.01	3	4	3	8	2	1	1	0	0	0
3	0.05	10	17	9	26	10	2	2	2	2	1
4	0.10	41	52	41	52	40	15	15	19	15	18
5	0.25	46	26	47	13	48	79	82	76	70	70
6	0.80	0	0	0	0	0	4	0	3	13	11
7	0.90	0	0	0	0	0	0	0	0	0	0
Proportion of patients treated (%)											
1	0.01	16	13	13	17	12	8	6	15	15	5
2	0.01	16	14	15	19	13	9	6	15	15	2
3	0.05	19	19	17	24	17	10	7	18	18	6
4	0.10	22	27	25	26	24	15	15	22	22	24
5	0.25	21	22	25	13	26	79	50	24	24	47
6	0.80	6	5	4	1	7	4	15	7	7	11
7	0.90	0	0	0	0	0	0	2	0	0	4
Average number of patients		20.0	24.3	23.9	32.2	17.1	12.5	18.3	21.0	21.0	21.0
Percent toxicity (%)		14.4	14.0	13.2	8.0	15.2	29.0	28.1	14.9	15.0	27.4

Table 6
Comparison of phase I designs under dose-toxicity Scenario 5.

Dose	DLT rate	Designs									
		3 + 3	3 + 3@9,2	CS(40; 3,6,9)	CS(25; 5,11)	CS(35; 2,6)	CS(50; 1,3,5)	CS(50)	mTPI	mTPI2	CRM
Proportion of recommended dose (MTD) dose levels (%)											
0		83	96	88	99	82	40	49	26	12	0
1	0.45	16	4	11	1	16	31	32	67	74	80
2	0.55	1	0	1	0	2	17	13	6	14	18
3	0.60	0	0	0	0	0	12	6	0	0	2
Proportion of patients treated (%)											
1	0.45	90	91	94	99	87	59	64	91	91	75
2	0.55	9	8	6	1	12	25	21	9	9	24
3	0.60	1	1	0	0	2	17	15	0	0	1
Average number of patients		4.7	5.1	5.0	3.3	5.1	7.0	12.5	5.7	5.7	6.0
Percent toxicity (%)		59.9	60.4	59.9	62.3	58.1	59.6	60.8	46.0	46.0	47.5

Table 7
Comparison of phase I designs under dose-toxicity Scenario 6.

Dose	DLT rate	Designs									
		3 + 3	3 + 3@9,2	CS(40; 3,6,9)	CS(25; 5,11)	CS(35; 2,6)	CS(50; 1,3,5)	CS(50)	mTPI	mTPI2	CRM
Proportion of recommended dose (MTD) dose levels (%)											
0		11	21	10	30	10	2	2	8	11	0
1	0.10	50	59	52	61	48	17	18	28	16	23
2	0.30	30	17	28	8	30	24	28	50	52	60
3	0.40	10	3	10	1	13	57	52	14	21	17
Proportion of patients treated (%)											
1	0.10	50	57	53	74	46	24	21	44	42	31
2	0.30	35	32	33	23	36	28	25	45	44	56
3	0.40	14	12	14	3	18	48	54	11	14	13
Average number of patients		10.9	14.5	14.4	14.2	10.7	8.4	14.8	11.9	11.8	12.0
Percent toxicity (%)		26.0	24.7	23.2	20.4	24.7	29.7	30.9	24.5	22.9	26.7

designs, but also treat more patients at or near the estimated MTD. Notably, both CS(50) and CS(50; 1,3,5) enroll fewer subjects and estimate the MTD as dose 17. CS(40; 3,6,9) has a low DLT proportion comparable to that of the 3 + 3 design, but treats a larger fraction of subjects near the estimated MTD. These advantages come at the expense of a larger sample size. We set the sample size for mTPI and mTPI2 to n = 18 to emulate the 3 + 3, to which they performed similarly. Setting the sample size for the CRM to be the same as for the 3 + 3, it yields the highest proportion of patients treated at or near the MTD.

Table 8

Application of the designs to data generated according to the trial in Simon et al. [7] CS(25) represents CS(25; 5,11), and CS(35) represents CS(35; 2,6).

	Design									
	3 + 3	3 + 3@9,2	CS(40; 3,6,9)	CS(25)	CS(35)	CS(50; 1,3,5)	CS(50)	mTPI	mTPI2	CRM
Patients enrolled	17	20	20	27	13	11	15	18	18	17
DLT rate at $\widehat{MTD}$ (%)	32	32	32	14	14	32	32	32	32	32
% with a DLT	18	15	15	11	23	18	40	17	17	29
% treated at $\widehat{MTD}$	35	45	45	41	46	45	67	33	33	76
% treated within 1 level of $\widehat{MTD}$	65	70	70	81	85	82	87	67	67	94

5. Discussion

We have proposed a family of rule-based phase I designs that retains the simplicity of the 3 + 3 while addressing its inflexibility and inefficiency. Unlike the 3 + 3, which targets DLT rates in the range 20%–25%, with our approach one can select a design to reflect any targeted DLT rate by means of the tuning parameter θ. Simulations suggest that choosing θ to be 10%–20% higher than the target toxicity probability gives the best chance of having the target dose be the modal dose, although this varies by scenario. The choice of J, the maximum number of cohort sizes, largely controls the total number of patients enrolled. If the sample size available for the trial is comparable to those typically in use with 3 + 3 designs, then $J = 2$ works well for lower target toxicity rates, whereas $J \in \{3, 4, 5\}$ works well for target toxicity rates of 25% or higher.

When the target toxicity is between 10% and 25%, CS(35; 2,6) is a practical choice. When the target toxicity exceeds 25%, we recommend CS(50; 1,3,5) for a smaller total sample size or CS(50; 1,3,5,8,10) when more patients are available. Using a large final cohort increases sample size requirements but eliminates the need for an add-on dose-expansion cohort.

The novelty of our design is that one can specify a sequence of cohort sizes, allowing for fast escalation at low doses and more enrollment close to the potential MTD. The possibility of specifying more than two cohort sizes sets our design apart from typical rule-based designs, where the cohort size is restricted to A + B or a multiple of a smaller cohort size.

The cohort-sequence design improves efficiency by escalating rapidly through the lower, safer doses and increasing cohort sizes adaptively when one begins to encounter toxicities. Consequently, it generally enrolls more patients in the vicinity of the estimated MTD and incurs higher overall DLT rates. Thus it avoids the wasteful assignment of subjects to doses that are likely to be safe and ineffective. In this way the cohort-sequence design paints a clearer picture of the drug's toxicity profile and increases the chance of clinical responses.

Our design effectively generalizes the 3 + 3; the CS(40; 3,6,9) version is comparable to the 3 + 3@9,2, which is a 3 + 3 with added dose-expansion cohort. Unsurprisingly, these two designs and the 3 + 3 perform comparably on most metrics, except that the 3 + 3 enrolls fewer patients because it lacks a built-in expansion cohort. The CS(35; 2,6) also performs similarly to the 3 + 3, with slightly fewer patients.

We have examined our design only in conventional scenarios where toxicity is nondecreasing with dose. Some contemporary cancer treatments, such as immunotherapy, can exhibit non-monotone dose-toxicity curves. In such cases, one may prefer a method that optimizes response rate subject to a maximum acceptable toxicity. Yet in a trial whose main aim is to study *safety*, the imperative is to identify, and pull back from, toxic doses. We therefore believe that there is a continuing role for traditional designs that operate on this principle, even in trials where a non-monotone toxicity curve is possible.

Although the critical values of the cohort-sequence design reflect Bayesian notions of parameter uncertainty, unlike the model-based designs it estimates the MTD based only on information from patients treated at the identified MTD. An alternative, hybrid approach that uses all the data would be to run the study with a rule-based design and then

estimate a model (such as the logistic) for the dose-response data, designating as MTD the dose whose predicted DLT rate is closest to, but does not exceed, the target rate.

In the last quarter-century, much statistical research has focused on model-based designs such as the CRM, in which one assumes an underlying parametric dose-response model and uses the accumulated data to estimate parameters and, thereby, the MTD [1,2,12,17–21]. Some pharmaceutical companies now use such designs routinely, often in "bucket" trials or phase I/II designs. The large majority of phase I cancer trials, however, continue to employ rule-based designs, primarily the 3 + 3 [13]. This reluctance to adopt the newer methods may reflect several factors: The cost and complexity of implementing the model-based methods; lack of familiarity with them; or simply a conviction that "better is the enemy of good enough" [22].

Model-based designs such as the CRM aim to identify the dose that delivers a target DLT rate [17]. By assuming an underlying dose-response model, all the data come into play at every decision point. These designs are accurate and efficient, provided only that the assumed model is roughly correct. Some implementations of model-based designs are resource-intensive, as they demand potentially a complete reanalysis of the data and re-evaluation of the dosing scheme at the time of enrollment of each new patient [12,22]. We emphasize that it is not the computations — which are now routine — that slow things down. The problem is the need to decide how to proceed at every interim analysis, which can easily consume more statistician and clinician time than many institutions are able and willing to allocate. That rule-based designs automate these decisions may explain their continued popularity.

References

[1] C. Le Tourneau, J.J. Lee, L.L. Siu, Dose escalation methods in phase I cancer clinical trials, J. Natl. Cancer Inst. 101 (2009) 708–720.

[2] T.M. Braun, The current design of oncology phase I clinical trials: progressing from algorithms to statistical models, Chin. Clin. Oncol. 3 (2014).

[3] Y. Ji, P. Liu, Y. Li, B. Nebiyou Bekele, A modified toxicity probability interval method for dose-finding trials, Clin. Trials 7 (2010) 653–663.

[4] G.M. Wheeler, M.J. Sweeting, A.P. Mander, AplusB: a web application for investigating A+B designs for phase I cancer clinical trials, PloS One 11 (2016), e0159026.

[5] Y. Ji, S.J. Wang, Modified toxicity probability interval design: a safer and more reliable method than the 3+3 design for practical phase I trials, J. Clin. Oncol. 31 (2013) 1785–1791.

[6] L. Edler, I. Burkholder, Overview of phase I trials, in: Crowley, Ankerst (Eds.), Handbook of Statistics in Clinical Oncology, second ed., Chapman & Hall/CRC, 2005, pp. 23–50.

[7] R. Simon, L. Rubinstein, S.G. Arbuck, M.C. Christian, B. Freidlin, J. Collins, Accelerated titration designs for phase I clinical trials in oncology, J. Natl. Cancer Inst. 89 (1997) 1138–1147.

[8] A. Ivanova, N. Flournoy, Y. Chung, Cumulative cohort design for dose-finding, J. Stat. Plann. Inference 137 (2007) 2316–2327.

[9] A.P. Oron, D. Azriel, P.D. Hoff, Dose-finding designs: the role of convergence properties, Int. J. Biostat. 7 (2011) 1–7.

[10] Y. Ji, Y. Li, B.N. Bekele, Dose-finding in phase I clinical trials based on toxicity probability intervals, Clin. Trials 4 (2007) 235–244.

[11] S. Liu, Y. Yuan, Bayesian optimal interval designs for phase I clinical trials, J. Roy. Stat. Soc.: Series C (Appl. Stat.) 64 (2015) 507–523.

[12] B. Huang, P. Bycott, E. Talukder, Novel dose-finding designs and considerations on practical implementations in oncology clinical trials, J. Biopharm. Stat. 27 (2017) 44–55.

[13] S.B. Love, S. Brown, C.J. Weir, C. Harbron, C. Yap, B. Gaschler-Markefski, J. Matcham, L. Caffrey, C. McKevitt, S. Clive, C. Craddock, Embracing model-based designs for dose-finding trials, Br. J. Canc. 117 (2017) 332–339.

[14] A. Ivanova, Escalation, group and A+B designs for dose-finding trials, Stat. Med. 25 (2006) 3668–3678.

[15] A.A. Mokdad, X.J. Xie, H. Zhu, D.E. Gerber, D.F. Heitjan, Statistical justification of expansion cohorts in phase 1 cancer trials, Cancer 124 (2018) 3339–3345.

[16] C. Ahn, An evaluation of phase I cancer clinical trial designs, Stat. Med. 17 (1998) 1537–1549.

[17] J. OàĂŽQuigley, M. Pepe, L. Fisher, Continual reassessment method: a practical design for phase 1 clinical trials in cancer, Biometrics 46 (1990) 33–48.

[18] J. Babb, A. Rogatko, S. Zacks, Cancer phase I clinical trials: efficient dose escalation with overdose control, Stat. Med. 17 (1998) 1103–1120.

[19] A. Rogatko, J.S. Babb, M. Tighiouart, F.R. Khuri, G. Hudes, New paradigm in dose-finding trials: patient-specific dosing and beyond phase I, Clin. Canc. Res. 11 (2005) 5342–5346.

[20] Y.K. Cheung, R. Chappell, Sequential designs for phase I clinical trials with late-onset toxicities, Biometrics 56 (2000) 1177–1182.

[21] P.F. Thall, S.J. Lee, Practical model-based dose-finding in phase I clinical trials: methods based on toxicity, Int. J. Gynecol. Canc. 13 (2003) 251–261.

[22] B. Neuenschwander, M. Branson, T. Gsponer, Critical aspects of the Bayesian approach to phase I cancer trials, Stat. Med. 27 (2008) 2420–2439.