Chapter 10

Initial Design

10.1 Introduction

In situations where the clinical investigators believe the MTD is among the higher doses but are not comfortable treating patients at a high dose without first testing its lower doses, one may adopt a two-stage CRM that starts a trial according to an initial sequence $\{x_{i,0}\}$ with $x_{1,0} = d_1$. This chapter discusses the characterization and the choice of the initial sequence. To be specific, Section 10.2 introduces some basic concepts about the ordering of dose sequences. Section 10.3 presents the calibration approach for $\{x_{i,0}\}$, along with numerical recommendations in some common clinical scenarios. Practical issues such as sample size considerations related to the use of a two-stage CRM will be addressed in Section 10.4. The calibration techniques will be illustrated in Section 10.5 in the context of the NeuSTART.

10.2 Ordering of Dose Sequences

A two-stage CRM assigns doses in accordance with a predetermined, nondecreasing dose sequence $\{x_{i,0}\}$ until the first toxicity is seen, that is, $x_i = x_{i,0}$ for $i \leq \min\{j : Y_j = 1\}$ where

$$x_{i,0} \leq x_{i+1,0}. \tag{10.1}$$

The primary appeal of a two-stage design is that it allows a low starting dose, which is often the lowest dose level, that is, $x_{1,0} = d_1$. On the other hand, the initial dose sequence should avoid treating many patients at low and inefficacious doses, for which the 3+3 algorithm is criticized. A reasonable strategy is to escalate quickly initially and slow down as the trial moves to higher doses [98]. Mathematically, it implies that the initial cohort sizes are monotone, that is,

$$m_{01} \leq m_{02} \leq \cdots \leq m_{0,K} \tag{10.2}$$

where

$$m_{0k} = \sum_{i=1}^{N} I(x_{i,0} = d_k)$$

is the *initial cohort size* at dose d_k according to the sequence $\{x_{i,0}\}$. Under (10.1), the initial cohort sizes m_{0k}s uniquely define the initial design. In this chapter, we will focus on initial dose sequences that satisfy (10.1) and (10.2).

A fast dose escalation scheme is often perceived to be aggressive, and a slow one conservative. Thus, a notion of relative dose escalation speed is in order:

Definition 10.1 (partial ordering). *Let* $\mathbf{x} = \{x_i\}$ *and* $\tilde{\mathbf{x}} = \{\tilde{x}_i\}$ *denote two sequences of doses. Dose escalation in the sequence* $\mathbf{x}$ *is said to be faster or greater than that in the sequence* $\tilde{\mathbf{x}}$*, denoted by* $\mathbf{x} > \tilde{\mathbf{x}}$ *if* $x_i \geq \tilde{x}_i$ *for all i, with the inequality being strict for some i.*

To illustrate, let $\mathbf{x_0}$ denote the initial design used in a two-stage CRM $\mathscr{D}_2$, and $\tilde{\mathbf{x}}$ denote a dose sequence generated by this design, that is, $\tilde{x}_i = \mathscr{D}_2(H_i)$. Hence, $\mathbf{x_0}$ is a predetermined fixed sequence, whereas $\tilde{\mathbf{x}}$ is a random sequence that depends on the toxicity outcomes in the trial. The notion of compatibility (Definition 5.1) can then be expressed as the condition under which $\mathbf{x_0} > \tilde{\mathbf{x}}$ with probability one. In other words, compatibility requires that the initial design represents the fastest (or, so to speak, most aggressive) dose escalation plan that is permissible with respect to the trial objective θ. This is a sensible requirement: the initial design is the dose escalation plan when no toxicity is observed; and when there are any observed toxic outcomes, it is undesirable that we should escalate faster than when there is none.

Another application of Definition 10.1 is to compare two initial dose sequences used in a two-stage CRM. To illustrate, the dose escalation plan for the NeuSTART was designed using a two-stage CRM with $N = 33$ patients and $K = 5$ test doses. The study initial design is given by (7.3), which can be equivalently represented by $m_{01} = m_{02} = 3$, $m_{03} = 6$, $m_{04} = 9$, and $m_{05} = 12$. Table 10.1 shows this NeuSTART initial sequence (bottom row) along with three other sequences that may be used for the trial. Since larger initial cohort sizes m_{0k}s correspond to slower escalation, it is obvious that the group-of-three design in Table 10.1 is faster than the group-of-four design and the NeuSTART sequence $\mathbf{x_0^{Neu}}$.

Table 10.1 Four examples of initial designs for a trial with $K = 5$ and $N = 33$

Design	m_{01}	m_{02}	m_{03}	m_{04}	m_{05}	$m_{+,4}$
Group-of-three, $\mathbf{x_0^{(3)}}$	3	3	3	3	21	12
Group-of-four, $\mathbf{x_0^{(4)}}$	4	4	4	4	17	16
Increasing cohort size, $\mathbf{x_0^{Inc}}$	2	4	6	8	13	20
NeuSTART, $\mathbf{x_0^{Neu}}$	3	3	6	9	12	21

Furthermore, we can verify that

$$\mathbf{x_0^{Inc}} > \mathbf{x_0^{Neu}}$$

because

$$x_{i,0}^{Inc} \geq x_{i,0}^{Neu} \text{ for all } i$$

with the inequality being strict when $i = 3, 21$; cf., Figure 10.1. It is equally interesting to note that although the group-of-three escalation is faster than group-of-four, there exists a sequence (i.e., $\mathbf{x_0^{Inc}}$) that does not show clear ordering with neither

of these. In other words, strict ordering does not necessarily exist between two sequences, and as such, the initial dose sequences constitute a partially ordered set.

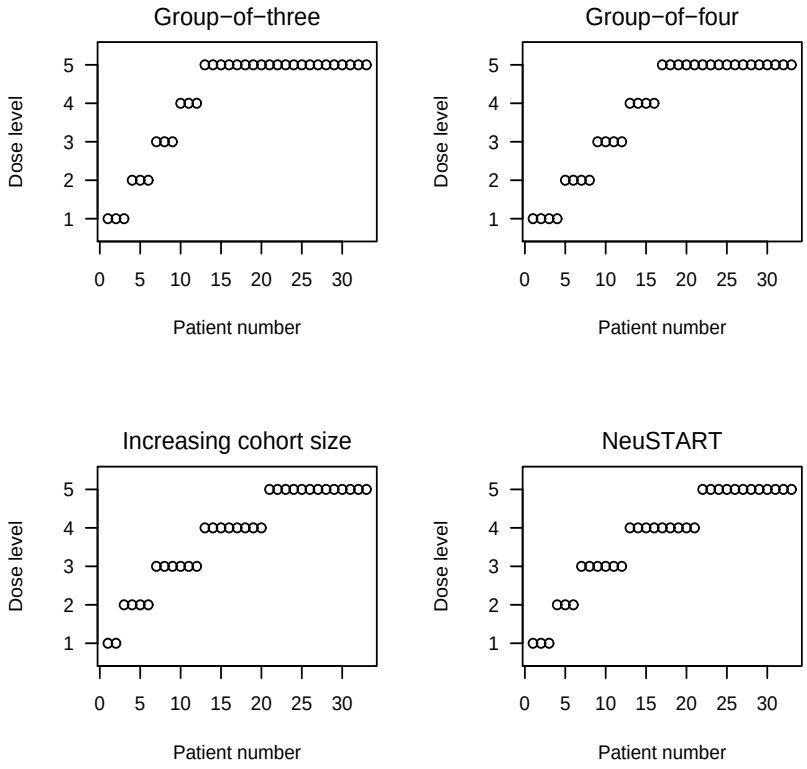

Figure 10.1 Dose level by patient number according to the initial designs in Table 10.1.

A simple index to indicate conservatism of an initial sequence is by the number of nontoxic observations required to reach the highest dose, that is,

$$m_{+,K-1} = \sum_{k=1}^{K-1} m_{0k}.$$

A conservative initial design is associated with a large value of $m_{+,K-1}$. In general, let the number of nontoxic observations required by an initial design to reach dose $j+1$ be denoted as

$$m_{+,j} = \sum_{k=1}^{j} m_{0k}. \tag{10.3}$$

Definition 10.2 (total ordering). *Let* $\mathbf{x} = \{x_i\}$ *and* $\tilde{\mathbf{x}} = \{\tilde{x}_i\}$ *denote two sequences of doses that satisfy (10.1) and (10.2), and* $\mathbf{m}$ *and* $\tilde{\mathbf{m}}$ *their respective cohort sizes. The sequence* $\mathbf{x}$ *is said to be more aggressive than* $\tilde{\mathbf{x}}$*, denoted by* $\mathbf{x} \succ \tilde{\mathbf{x}}$*, if* $m_{+,J_0} < \tilde{m}_{+,J_0}$ *where* $J_0 = \max\{j : m_{+,j} \neq \tilde{m}_{+,j}\}$.

A practical advantage of the relation " $\succ$ " is that it is a total order. As a result, we can use it to rank dose sequences according to their aggressiveness. For example, in Table 10.1, the sequence $\mathbf{x}_{\mathbf{0}}^{\mathbf{Neu}}$ is the least aggressive as it takes $m_{+,4} = 21$ consecutive nontoxic outcomes to escalate to dose level 5, whereas the group-of-three is the most aggressive with $m_{+,4} = 12$. Overall, we have

$$\mathbf{x}_{\mathbf{0}}^{(\mathbf{3})} \succ \mathbf{x}_{\mathbf{0}}^{(\mathbf{4})} \succ \mathbf{x}_{\mathbf{0}}^{\mathbf{Inc}} \succ \mathbf{x}_{\mathbf{0}}^{\mathbf{Neu}}.$$

10.3 Building Reference Initial Designs

10.3.1 Coherence-Based Criterion

As stated in Chapter 5, Section 5.2, a one-stage CRM is coherent, but a two-stage CRM is not necessarily so. A logical deduction is that the incoherence problem arises from improper choice of the initial design. Theorem 5.2 precisely indicates that a cause of incoherence is the use of an incompatible initial design. To see how incoherence may occur, consider an outcome sequence by a two-stage CRM with $x_{1,0} = \cdots x_{m',0} = d_1$ and $Y_1 = \cdots = Y_{m'-1} = 0$ and $Y_{m'} = 1$ such that $\mathscr{D}_2(H_{m'+1}) \geq d_2$. This sequence is incoherent because an escalation takes place for patient $m'+1$ after a toxic outcome is observed in patient m'. Indeed, this outcome sequence is made possible only when m_{01} is set to be sufficiently large, that is, $m_{01} \geq m'$, and can be avoided by choosing a small m_{01}. This illustrates that an overconservative initial design causes incoherence and incompatibility.

To appeal to our intuition, take another instance with a target $\theta = 0.25$. It is quite clear then that an initial design that escalates after every 10 nontoxic outcomes, that is, $m_{0k} \equiv 10$, is overconservative. Likewise, an initial sequence with $m_{0k} \equiv 8$ also appears conservative, though not as much as $m_{0k} \equiv 10$. As we decrease m_{0k}, the initial design seems to be increasingly reflective of the objective $\theta = 0.25$. The natural question is precisely how small m_{0k} should be so that the initial design is appropriate. This is the question the coherence (compatibility) criterion may address.

Example 10.1. Consider the two-stage CRM used in Chapter 3, Figure 3.1. The right panel of the figure shows that the group-of-three initial rule is incompatible. In Chapter 5, Section 5.2.2 subsequently illustrates that the group-of-two initial design is compatible; and so is the group-of-one rule. That is, among all initial designs with constant initial cohort sizes, that is, $m_{0k} \equiv m_0$, the only compatible choices are $m_{0k} = 1$ or 2. In view of the inclination to be conservative, one may adopt $m_{0k} = 2$ for $k = 1, \ldots, 4$ in conjunction with this CRM model.

10.3.2 Calibrating Compatible Dose Sequences

As compatibility and the inclination to be conservative are two opposing criteria, it is reasonable to calibrate the initial design to be the "most conservative" compatible design. There are then two possible approaches to implement the calibration process. The first approach starts with a compatible initial design, iterates to a less aggressive dose sequence at each step, and stops iteration once an incompatible initial design is attained. This calibration approach can be implemented by the following algorithm:

Algorithm 10.1. Base-b compatible benchmark

1. For given θ, K, ν_0, CRM model $F_k(\beta)$, and the prior on β, set $j = \nu_0$ and specify a base (b) initial design $m_{0k} = 0$ for $k < j$, and $= b$ for $k \geq j$.
2. Check the compatibility of the initial design $\{m_{01}, \ldots, m_{0,K-1}, m_{0K}\}^{\dagger}$.
3. If compatibility holds:
   (a) Record this initial design
   (b) Iterate
   $$j = \begin{cases} j-1 & \text{if } j > 1 \\ K-1 & \text{if } j = 1 \end{cases}$$
   and then update m_{0j} with $m_{0j} + b$, and repeat Step (2).
4. If compatibility fails to hold in Step 2, stop iteration and choose the last recorded initial design in Step 3a.

† The cohort size m_{0K} of the highest dose does not affect whether the initial design is compatible, and can be arbitrarily specified and updated in the algorithm as long as (10.2) is satisfied.

At each iteration, Step 3b moves to test a more conservative initial design. It can be proved that once an incompatible sequence is reached in Step 4, all subsequent dose sequences will be incompatible; see Theorem 10.1 in Section 10.4. Thus, the last recorded design in Step 3a may be viewed as a conservative benchmark for compatible dose sequences, and shall be called a base-b benchmark. It is easy to see any base-b benchmark satisfies the constraint (10.2).

Example 10.2. For the CRM model in Example 10.1, if we apply Algorithm 10.1 with $b = 1$, we will test dose sequences in the following order:

Iteration	m_{01}	m_{02}	m_{03}	m_{04}	Compatible
1	0	0	1	1	Yes
2	0	1	1	1	Yes
⋮	⋮	⋮	⋮	⋮	⋮
9	2	2	3	3	Yes
10	2	3	3	3	No

We stop at the tenth iteration when we reach an incompatible initial design. The dose

sequence in the ninth iteration, $m_{01} = m_{02} = 2$ and $m_{03} = m_{04} = 3$, is the base-1 benchmark.

At each iteration, compatibility can be verified using the function `cohere`. For instance, for the initial design at the ninth iteration,

```
> theta <- 0.25
> p0 <- c(0.05,0.12, 0.25, 0.40, 0.55)
> m0 <- c(2,2,3,3,3) # set m0[5]=3 arbitrarily
> x0 <- rep(1:5, m0)
> foo <- cohere(p0,theta,x0,model="logistic")
> foo$message
[1] "Coherent"
>
```

The base-1 benchmark in Example 10.2 is more conservative than the group-of-two design obtained in Example 10.1, and may agree more with clinicians' expectations. Ideally, one should seek a compatible initial sequence $\tilde{\mathbf{x}}_\mathbf{0}$ such that $\mathbf{x_0} > \tilde{\mathbf{x}}_\mathbf{0}$ for all compatible $\mathbf{x_0}$. Unfortunately, it is unclear whether such a sequence exists in general, mainly because of the fact that dose sequences form a partially ordered set. Table 10.2 displays all possible base-b benchmarks for this CRM model; when $b \geq 7$, there is no compatible initial sequence that starts at $x_{1,0} < d_K$. However, by extensive numerical search, the base-1 design appears to be the most conservative among all base-b designs. Since the starting dose is a clinical decision, there may be further restrictions on $x_{1,0}$. In particular, it is common to adhere to sequences with $x_{1,0} = d_1$, which implies $m_{01} > 0$. In this case, we will choose between the base-1 and base-2 sequences as the initial design for this given CRM model. As a two-stage CRM is motivated by conservatism, it is natural to choose base-1 benchmark in this example.

Table 10.2 Base-b initial sequences for the CRM model in Example 10.1

b	m_{01}	m_{02}	m_{03}	m_{04}	$m_{+,2}$	$m_{+,3}$	$m_{+,4}$
1	2	2	3	3	4	7	10
2	2	2	2	2	4	6	8
3	0	3	3	3	3	6	9
4	0	0	0	4	0	0	4
5	0	0	0	5	0	0	5
6	0	0	0	6	0	0	6

A second calibration approach proceeds in the opposite direction: start with an incompatible initial design, and increase escalation speed at each iteration until a compatible initial design is obtained. An analog of Algorithm 10.1 can be used to implement this calibration approach. However, it can be verified using Theorem 10.1 that this approach will yield the same base-b benchmark given by Algorithm 10.1.

10.3.3 Reference Initial Designs for the Logistic Model

This subsection presents some conservative references of initial designs that are to be used with the logistic model (3.3):

$$F(d_k, \beta) = \frac{\exp\{3 + \exp(\beta) d_k\}}{1 + \exp\{3 + \exp(\beta) d_k\}} \tag{10.4}$$

in a two-stage Bayesian CRM, where $\beta \sim N(0, 1.34)$ a priori and the dose labels d_ks are determined via a specified half-width δ by Algorithm 8.1. We take the following calibration approach: for a given CRM model, apply Algorithm 10.1 to obtain the base-b benchmark for each possible b with $m_{01} > 0$, and among them select the most conservative sequence according to the total order "$\succ$". (This is essentially what we did in Section 10.3.2.) The results are shown in Table 10.3 for the clinical parameters with

- $\theta = 0.10, 0.20, 0.25, 0.33$
- $K = 4, 5, 6, 7$.

The prior MTD ν_0 is set at the median dose level (7.2). The ranges of δ in the table are chosen in accordance with the optimal δ given in Tables 8.2–8.5 for cases with $N \geq 30$ and ν_0 at the median dose; the results here are, therefore, to be used in conjunction with those tables.

Table 10.3 demonstrates some general trends regarding the escalation speed of the reference initial designs. First of all, all reference designs turn out to be base-1. Furthermore, the initial dose sequence will need to be more aggressive as

- A higher target θ is used, or,
- A larger number K of doses are tested, or,
- The CRM model is specified by a smaller half-width δ.

The first trend is expected; the second trend is also quite clear. The third trend is comparatively nuanced, but provides an additional consideration for the choice of δ in the context of a two-stage CRM. For example, one may be inclined to choose a δ that corresponds to a more conservative reference initial design.

10.4 Practical Issues

10.4.1 Sample Size Constraint

Algorithm 10.1 prescribes a base-b benchmark without regard for the initial cohort size m_{0K} at the highest dose. In practice, it may be prudent to set aside an adequate number m^* of observations at the highest level in case of no toxicity, that is,

$$m_{0K} \geq m^*.$$

The choice of m^* can be partly informed by the target rate θ. For example, if we set $m^* = 1/\theta$, we will expect to observe one toxic outcome at the highest dose if no toxic outcome is observed at the lower doses. To account for the possibility that

Table 10.3 Reference initial designs for model (10.4). All designs are base-1.

θ	K	δ	m_{01}	m_{02}	m_{03}	m_{04}	m_{05}	m_{06}
0.10	4	.03	6	6	7	—	—	—
		.04	8	9	9	—	—	—
	5	.02	4	4	4	4	—	—
		.03	6	6	6	6	—	—
	6	.02	4	4	4	4	4	—
	7	.02	3	3	3	4	4	4
0.20	4	.06	3	3	4	—	—	—
	5	.05	2	3	3	3	—	—
		.06	3	3	3	4	—	—
	6	.05	2	2	3	3	3	—
	7	.04	2	2	2	2	2	2
		.05	2	2	2	2	3	3
0.25	4	.06	2	2	3	—	—	—
		.07	3	3	3	—	—	—
	5	.05	2	2	2	2	—	—
		.06	2	2	2	2	—	—
		.07	2	2	3	3	—	—
	6	.05	1	2	2	2	2	—
		.06	2	2	2	2	2	—
	7	.05	1	1	2	2	2	2
		.07	2	2	2	2	3	3
0.33	4	.06	1	2	2	—	—	—
		.07	2	2	2	—	—	—
		.09	2	2	2	—	—	—
	5	.06	1	1	1	2	—	—
		.07	1	1	2	2	—	—
		.09	2	2	2	2	—	—
	6	.06	1	1	1	1	2	—
		.07	1	1	1	2	2	—
	7	.06	1	1	1	1	2	2
		.07	1	1	1	1	2	2

the actual escalation is slower and more patients will be treated at the lower doses, we may choose $m^* \geq \lambda/\theta$ for some $\lambda > 1$. As a practical guideline, we recommend choosing λ to be somewhere between 1.5 and 2.0. This recommendation serves as a lower limit of the required sample size; but the final choice should be evaluated based on the operating characteristics by simulations. At any rate, the sample size N is determined by the choice of m^* and the initial design:

$$N \geq m_{+,K-1} + m^*. \tag{10.5}$$

For example, if we use the reference initial design corresponding to $\theta = 0.25$, $K = 5$, and $\delta = 0.07$ in Table 10.3, and set $m^* = 1.5/0.25 = 6$, then we may use $N \geq 16$. That is, the constraint (10.5) can be used to give a quick assessment as to whether the sample size N is adequate; cf,. constraint (7.1) for the one-stage CRM.

In many practical situations, however, the sample size N is limited by the time and resources available; and a base-b benchmark may require more patients than the available resources can provide. Consider the NeuSTART where $N = 33$, for instance. If we use the reference initial design corresponding to $\theta = 0.10$, $K = 5$, and $\delta = 0.03$ in Table 10.3, the trial will have enrolled 24 subjects before reaching the highest dose and left $m_{05} = 9$ for the dose—if no toxicity occurred in the first 24 subjects; the trial may not reach dose level 5 with this initial design even with one observed toxic outcome. A pruned base-b benchmark may then be used:

Algorithm 10.2. Pruning a compatible sequence for given N and m^*

1. For given θ, K, ν_0, $F_k(\beta)$, the prior on β, specify a compatible initial design $\{m_{01}, \ldots, m_{0,K-1}\}$ and set $m_{0K} = N - m_{+,K-1}$.
2. If $m_{0K} < m^*$, set $j = 0$ and proceed with the following steps:
   (a) Iterate
   $$j = \begin{cases} j+1 & \text{if } j < K-1 \\ 1 & \text{if } j = K-1 \end{cases}$$
   and then update m_{0j} with $m_{0j} - 1$ and m_{0K} with $m_{0K} + 1$.
   (b) Stop iteration when $m_{0K} = m^*$.

Example 10.3. Consider $N = 33$ and $m^* = 12$ in the context of the NeuSTART. The initial design $m_{01} = m_{02} = 5$, $m_{03} = m_{04} = 10$ in Table 10.3 may be pruned according to Algorithm 10.2 in the following order:

Iteration	m_{01}	m_{02}	m_{03}	m_{04}	m_{05}
0	6	6	6	6	9
1	5	6	6	6	10
2	5	5	6	6	11
3	5	5	5	6	12

The algorithm stops at the ninth iteration, where a compatible initial design with $m_{05} = 12$ is attained.

Step 2a of Algorithm 10.2 reverses Algorithm 10.1 and yields a faster initial sequence at each iteration; and hence the pruned sequence will be less conservative than the base-b benchmark. As a consequence of Theorem 10.1 shown later, the pruned sequence will also be compatible.

Pruning may sometimes lead to an unacceptably aggressive initial design. Should this be the case, an appropriate approach is to increase the sample size N. That is, we are to use (10.5) as a sample size constraint. In practice, it may be useful to produce several pruned compatible designs for a range of Ns and a given m^*, solicit the most

feasible design from a clinical perspective, and confirm the design's performance by simulation. In Example 10.3, if $N = 36$ is a feasible sample size, the unpruned initial design with $m_{01} = m_{02} = m_{03} = m_{04} = 6$, and $m_{05} = 12$ may also be presented along with the pruned sequence, so that the clinical investigator may weigh in with the tangible trade-off between sample size and conservatism.

10.4.2 Dose Insertion

Dose insertion in phase I trials is a common idea among practitioners. Take the NeuSTART, for example. The investigators had suggested inserting an intermediate dose at 7 mg/kg/day between dose tier 3 (6 mg/kg/day) and dose tier 4 (8 mg/kg/day) in case toxicities are observed at dose tier 4. The motivation was to test a dose higher than 6 mg/kg/day (dose tier 3) if a deescalation from dose tier 4 was needed as a result. (This suggestion was not implemented in the study from a pharmacological perspective; cf., Section 7.2.2.)

From a statistical viewpoint, the cleanest way to handle dose insertion is to avoid any post hoc addition. That is, the suggested intermediate dose or doses should be included for potential testing in the planning stage. In the NeuSTART, if the dose 7 mg/kg/day is to be used, then one should plan a trial with $K = 6$ dose levels; the revised dose tiers will then be

Dose tier, k	1	2	3	4	5	6
Lovastatin dose (mg/kg/day)	1	3	6	7	8	10

To reflect that 7 mg/kg/day (dose tier 4) is inserted, we may prescribe an initial design that skips the dose in a two-stage CRM, that is, the initial cohort size $m_{04} = 0$. The calibration of the initial design can go through the similar process via Algorithms 10.1 and 10.2 otherwise.

A simpler alternative is to use the reference initial design in Table 10.3. For the revised dose tiers for NeuSTART, we may use the initial design corresponding to the CRM model for $\theta = 0.10, K = 6$, and $\delta = 0.02$ from the table, but set $m_{04} = 0$, that is,

$$m_{01} = m_{02} = m_{03} = 4, m_{04} = 0, m_{05} = 4. \tag{10.6}$$

Theorem 10.1. *Let* $\mathbf{m_0} = \{m_{0k}\}$ *and* $\tilde{\mathbf{m}}_\mathbf{0} = \{\tilde{m}_{0k}\}$ *denote the initial cohort sizes of two sequences* $\mathbf{x_0}$ *and* $\tilde{\mathbf{x}}_\mathbf{0}$*, respectively, where* $m_{0k} \leq \tilde{m}_{0k}$ *for all k and (10.1) and (10.2) are satisfied. If* $\tilde{\mathbf{x}}_\mathbf{0}$ *is a compatible initial design for a given CRM model, then* $\mathbf{x_0}$ *is also compatible.*

Theorem 10.1 guarantees that (10.6) is a compatible initial design because the reference design obtained from Table 10.3 is compatible.

10.5 Case Study: NeuSTART

The NeuSTART adopted a two-stage CRM with $\theta = 0.10, K = 5$, and $\nu_0 = 3$. The empiric (3.2) function

$$F(d_k, \beta) = d_k^{\exp(\beta)} \tag{10.7}$$

was used to model the toxicity probability at dose level k with labels $d_1 = 0.02$, $d_2 = 0.06$, $d_3 = 0.10, d_4 = 0.18$, and $d_5 = 0.30$, where $\beta \sim N(0, 1.34)$ a priori. The initial dose sequence is given in (7.3). The calibration process of this CRM design is described in Section 7.4.2, and is characterized as a trial-and-error approach. In the following, we calibrate the initial design by Algorithms 10.1 and 10.2 for this study model (10.7). For comparison purposes, we set $N = 33$ and $m^* = 12$ (although in retrospect, we should have set $m^* = 15$). First, Table 10.4 displays the base-b benchmarks for all possible b so that $m_{01} > 0$. Of all sequences, the base-1 design is most conservative according to the total order $\succ$. Next, pruning the base-1 design with $N = 33$ and $m^* = 12$, we obtain

$$m_{01} = 4, m_{02} = 5, m_{03} = m_{04} = 6, \text{ and } m_{05} = 12. \tag{10.8}$$

It can be easily verified that the pruned design (10.8) is slower than the original initial sequence (7.3).

Table 10.4 Base-b initial sequences for the NeuSTART model

b	m_{01}	m_{02}	m_{03}	m_{04}	$m_{+,2}$	$m_{+,3}$	$m_{+,4}$
1	7	7	8	8	14	22	30
2	6	6	8	8	12	20	28
3	6	6	9	9	12	21	30
4	4	8	8	8	12	20	28
5	5	5	10	10	10	20	30
6	6	6	6	12	12	18	30
7	7	7	7	7	14	21	28

Suppose we opt to adopt the logistic model (10.4) instead of the empiric model in NeuSTART. We may choose $\delta = 0.02$ or 0.03 as recommended in Table 10.3. Further suppose we use $\delta = 0.03$. Then the pruned initial sequence is given in Example 10.3, as follows:

$$m_{01} = m_{02} = m_{03} = 5, m_{04} = 6, \text{ and } m_{05} = 12. \tag{10.9}$$

This initial dose sequence is more conservative than (10.8) and the original (7.3). For comparison, Table 10.5 shows the simulation results of two-stage CRM using these three initial dose sequences, along with that of the nonparametric optimal design. The three CRM designs have comparable accuracy—all three have average PCS 0.63—but the logistic model with (10.9) has more varied PCS than the other two. In comparison, the average PCS of the nonparametric optimal design is 0.76.

The three CRM designs have comparable toxicity and overdose. Importantly, all three lead to much reduced OD numbers when compared to balanced randomization. This illustrates the conservative nature of the two-stage CRM design.

As we have emphasized in the previous chapters, the CRM designs obtained by the automated calibration approach give comparable operating characteristics to the original labor-intensive NeuSTART design. In particular, Table 10.3 provides a quick start to calibrate and prune a reasonable initial design for a two-stage CRM.

Table 10.5 The distribution of dose selection, average toxicity number (ATN), and overdose number (OD) of three two-stage CRM for the NeuSTART with $\theta = 0.25, K = 5, \nu_0 = 3, N = 33$, and $m^* = 12$

Model	$\mathbf{x_0}$	Probability of selecting dose					ATN	OD
Scenario 1:		**.10**	.25	.30	.35	.40		
NP optimal	—	**.92**	.07	.01	.00	.00	—	—
Empiric	(7.3)	**.88**	.11	.01	.00	.00	4.6	7.6
Empiric	(10.8)	**.88**	.11	.01	.00	.00	4.5	7.1
Logistic	(10.9)	**.86**	.13	.02	.00	.00	4.5	7.2
Scenario 2:		.04	**.10**	.25	.30	.35		
NP optimal	—	.23	**.70**	.07	.01	.00	—	—
Empiric	(7.3)	.32	**.53**	.14	.01	.00	3.6	6.8
Empiric	(10.8)	.32	**.52**	.15	.01	.00	3.5	6.2
Logistic	(10.9)	.29	**.53**	.15	.02	.01	3.6	6.5
Scenario 3:		.01	.04	**.10**	.25	.30		
NP optimal	—	.01	.22	**.70**	.07	.01	—	—
Empiric	(7.3)	.02	.27	**.56**	.14	.01	3.1	5.4
Empiric	(10.8)	.02	.26	**.56**	.13	.02	3.0	5.1
Logistic	(10.9)	.02	.25	**.54**	.15	.04	3.1	5.8
Scenario 4:		.01	.01	.04	**.10**	.25		
NP optimal	—	.00	.01	.22	**.70**	.08	—	—
Empiric	(7.3)	.00	.03	.25	**.56**	.16	2.4	2.9
Empiric	(10.8)	.00	.03	.26	**.51**	.19	2.3	3.3
Logistic	(10.9)	.00	.03	.24	**.46**	.27	2.4	4.3
Scenario 5:		.01	.01	.01	.04	**.10**		
NP optimal	—	.00	.00	.01	.22	**.77**	—	—
Empiric	(7.3)	.00	.01	.05	.28	**.66**	1.4	0.0
Empiric	(10.8)	.00	.01	.06	.26	**.67**	1.3	0.0
Logistic	(10.9)	.00	.01	.04	.19	**.76**	1.3	0.0

Note: The results for the nonparametric optimal design (NP optimal) are given as references. Numbers associated with the MTD are in bold.

10.6 Exercises and Further Results

Exercise 10.1. Verify that $\mathbf{x} > \tilde{\mathbf{x}}$ implies $\mathbf{x} \succ \tilde{\mathbf{x}}$.

Exercise 10.2. Redesign the NeuSTART as in Section 10.5. For the logistic model (10.4) with $\delta = 0.02$, give an initial sequence under the constraint $N = 33$ and $m^* = 12$. Use simulation to compare this two-stage CRM with the designs in Table 10.5.

Exercise 10.3. Using Theorem 10.1, verify that pruning using Algorithm 10.2 will result in a compatible sequence.