CPRC Initial Full Review Application

Appendix A: Phase I Trial Operating Characteristics

# Overview

The purpose of this appendix is to document the operating characteristics of Phase I trials with dose-escalation designs. If operating characteristics are described in the protocol, this appendix is not required and the application submitter should indicate the location of the description of operating characteristics in the CPRC application (section 4).  
  
To evaluate the performance of the dose escalation design, CPRC biostatistical reviewers require documentation of operating characteristics for the study design under one or more sets of hypothesized toxicity proportions associated with each dose level, in order to determine how likely it is that this design will select a dose in the desired toxicity range. **The CPRC is not requiring revision to the protocol document to include this information.**

# Background

*What are operating characteristics and why are they useful?*In a typical phase I trial, the primary goal is to select from a sequence of dose levels the maximum tolerated dose (MTD), which is the highest dose level for which the proportion of patients who experience dose-limiting toxicity (DLT) is at or below a predefined threshold. Another goal is to treat as many patients as possible near the MTD, and avoid treating many patients at dose levels that are too toxic.  
  
Operating characteristics (OCs) evaluate how well a clinical trial design would meet its goals given some assumptions about the effects of the treatment on a patient population. One can think of OCs as a simulated stress test to determine whether a design is likely to produce reliable results.   
  
OCs in a phase I trial serve an analogous purpose to power analyses, which are standard in later phase trials. In a power analysis, researchers make assumptions about treatment effects, and then calculate the probability that their trial design will produce the desired result given those assumptions. For example, in a two-arm trial with a binary endpoint, one might assume that the true response proportion in the control group is 30%, and the true response proportion in the treatment group is 50%. Even if this assumption is true, in the actual trial, the response proportion in each group may be higher or lower than the true proportions, and there may or may not appear to be a difference in response proportion for treatment versus control. A power analysis gives the probability (power) that at the end of the trial, the response proportion is statistically significantly higher in the treatment group (given the assumptions above, plus a sample size, statistical model or test, and decision criterion). If power is lower than desired, then one should consider increasing the sample size or otherwise modifying the design. It would be shortsighted to proceed with a phase II or III trial without conducting a power analysis.   
  
In a phase I trial, instead of directly comparing treatment groups, one is attempting to select the MTD. Analogous to a power analysis, OCs tell us the likelihood that a design will actually select the optimal MTD.  
  
In a dose escalation trial, prior to calculating OCs, the following parameters must be defined: *the number of dose levels, dosages, and maximum acceptable DLT proportion for the MTD, such as 10% or 30%*. Researchers then make assumptions regarding the true proportion of patients who will have DLT at each dose level. For example, in a trial with three dose levels, one might hypothesize that the true DLT proportions are 10%, 25%, and 40%. Then, the OCs can be calculated to demonstrate how the design will perform if the true DLT proportions at the three dose levels are 10%, 25% and 40%. Of course, these are guesses, which is why it may be beneficial to calculate OCs for multiple clinically plausible scenarios.  
  
There are three important variables that should be included for each dose level in OC output:

a. *Probability that each dose level will be chosen as the MTD (and the probability that no dose will be selected because they are all too toxic)*. If the optimal dose level is not selected often enough, one could consider a different design or maximum sample size, or anticipate that a dose expansion cohort could be necessary to confirm the MTD. Also, some phase I designs are inflexible with respect to defining the MTD DLT proportion. The 3+3 design tries to select dose levels associated with 20-30% of patients having DLT. If your desired MTD DLT proportion is 10% or less, another design is probably more appropriate.

b. *Expected number of patients treated at each dose level*. Some designs are better than others at treating more patients at dose levels near the MTD under certain scenarios. If low toxicity is expected, some designs move through lower dose levels faster.

c. *Expected number of DLTs at each dose level.* This is useful for evaluating whether the design is good at rejecting dose levels that are too toxic before treating too many patients at those levels, especially if toxicity is higher than expected.

There are multiple online OC calculators for common phase I trial designs, some of which are referenced below. However, to be useful, one has to carefully consider the DLT scenarios, and also interpret the OC output with attention to trade-offs.   
  
The purpose of OCs is to help researchers choose better trial designs and to understand the limitations of those designs. It is hard to determine the true MTD when treating only a few patients at each dose level. When we have more information about how a design will perform, we can hopefully make better trial design decisions which increase the likelihood that more patients will be treated at optimal dose levels, both within the phase I trial and in subsequent trials that use the same treatment.

# Software

Examples of programs that can be used to calculate operating characteristics are listed below. They are provided as references only; other programs and/or designs may be used.

* 1. Trialdesign.org calculates operating characteristics for several common dose escalation designs including BOIN, CRM, and Keyboard. This is usually found on the simulation tab for each design. The BOIN program has an option to calculate 3+3 operating characteristics (<https://www.trialdesign.org/one-page-shell.html#BOIN>).
  2. The R package “UBCRM” has functions ssimCrm and ssim3p3 that calculate operating characteristics for the CRM and 3+3 designs, respectively.

# Example

*The following text represents an excellent example of a response to this documentation request.*

The operating characteristics of the BOIN design were assessed via clinical trial simulations under various assumed true dose-toxicity relationships. Simulations were performed under 5 different dose-toxicity scenarios. Table 2 gives the true DLT probability at each study drug dose level under each simulation scenario.

The scenarios represent a range of locations for the true MTD. For each scenario, the DLT probability for the true MTD is bolded. Note that all dose levels exceed the true MTD for scenario 1.

**Table 2.** True DLT Probability at Each Dose Level for Simulation Scenarios

|  |  |  |  |
| --- | --- | --- | --- |
|  | Study Drug Dose (mg) | | |
| Scenario | **200** | **400** | **800** |
| **1** | 0.480 | 0.600 | 0.700 |
| **2** | **0.285** | 0.480 | 0.600 |
| **3** | 0.085 | **0.285** | 0.500 |
| **4** | 0.050 | 0.100 | **0.285** |
| **5** | 0.050 | 0.100 | **0.150** |

To assess the operating characteristics of the BOIN design, 1000 trials were simulated for each of the 5 scenarios shown in Table 2. For simplicity, the following conventions were implemented for all simulations:

* The initial cohort assigned to a dose level consisted of exactly 6 subjects, and subsequent cohorts (if any) assigned to the same dose level consisted of exactly 3 subjects
* The maximum sample size was 24 subjects
* If the starting dose level (200 mg) was eliminated by the dose escalation rule, the trial was terminated early for excessive toxicity (and no MTD was selected)
* If the dose escalation rule indicated allocation to a dose level for which 9 subjects had previously been allocated, the trial was stopped to declare the MTD (i.e., the maximum sample size at any dose level was 9 subjects)

If the lowest dose level did not meet the MTD criteria at the conclusion of the trial, no MTD was selected (i.e., the lowest dose level required de-escalation per the BOIN decision rule, such that all dose levels exceeded the MTD.)

The simulated trials were summarized to obtain the following operating characteristics of the BOIN design under each dose-toxicity scenario:

* MTD selection percentage (% of trials which select each dose level as the MTD)
* Average number of subjects allocated per dose level
* Average total number of subjects allocated per trial
* Average number of DLTs observed per dose level
* Average total number of DLTs observed per trial

Tables 3 through 7 display the operating characteristics for each simulation scenario. The results indicate that the BOIN design correctly selects the true MTD more frequently than other dose levels, and generally selects a dose level at or near the true MTD. Note that for scenario 1, the BOIN design correctly selects no MTD (i.e., all doses exceed the true MTD) with high probability.

**Table 3.** Operating Characteristics for Scenario 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dose (mg) | True Prob (DLT) | MTD Selection % | Avg N | Avg DLT |
| **200** | 0.480 | 12.5 | 8.1 | 3.8 |
| **400** | 0.600 | 0.3 | 1.2 | 0.7 |
| **800** | 0.700 | 0.0 | 0.1 | 0.1 |
|  |  | No MTD = 87.2% | 9.4 | 4.6 |

**Table 4.** Operating Characteristics for Scenario 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dose (mg) | True Prob (DLT) | MTD Selection % | Avg N | Avg DLT |
| **200** | 0.285 | 48.2 | 8.7 | 2.4 |
| **400** | 0.480 | 6.6 | 4.5 | 2.1 |
| **800** | 0.600 | 0.1 | 0.6 | 0.4 |
|  |  | No MTD = 45.1% | 13.8 | 4.9 |

**Table 5.** Operating Characteristics for Scenario 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dose (mg) | True Prob (DLT) | MTD Selection % | Avg N | Avg DLT |
| **200** | 0.085 | 44.6 | 7.8 | 0.6 |
| **400** | 0.285 | 46.8 | 8.4 | 2.3 |
| **800** | 0.500 | 5.5 | 4.3 | 2.2 |
|  |  | No MTD = 3.1% | 20.5 | 5.2 |

**Table 6.** Operating Characteristics for Scenario 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dose (mg) | True Prob (DLT) | MTD Selection % | Avg N | Avg DLT |
| **200** | 0.050 | 5.1 | 6.5 | 0.3 |
| **400** | 0.100 | 46.9 | 7.9 | 0.8 |
| **800** | 0.285 | 47.5 | 8.1 | 2.3 |
|  |  | No MTD =0.5% | 22.5 | 3.5 |

**Table 7.** Operating Characteristics for Scenario 5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dose (mg) | True Prob (DLT) | MTD Selection % | Avg N | Avg DLT |
| **200** | 0.050 | 5.0 | 6.5 | 0.3 |
| **400** | 0.100 | 12.9 | 6.9 | 0.7 |
| **800** | 0.150 | 81.6 | 8.3 | 1.2 |
|  |  | No MTD =0.5% | 21.6 | 2.3 |