# From Tradition to Innovation: Assessing the Efficiency of Dose-Escalation Methods in Phase I Oncology Trials

**Megan Wang** [1*]

[1]Bellaire High School, Bellaire, TX, USA
*Corresponding Author: meganwang28@gmail.com

**Advisor: Dr. Suyu Liu,**
The University of Texas MD Anderson Cancer Center, Houston, TX, USA, syliu@mdanderson.org

**Abstract**

Phase I clinical trials are essential in developing anticancer drugs, serving as the first step in testing new treatments in humans. Their primary goal is to determine the maximum-tolerated dose (MTD) - the highest dose patients can receive with tolerable side effects. The traditional 3+3 design has dominated phase I trials for decades due to its simplicity, but it is often criticized for its poor performance in identifying the true MTD and for assigning patients to suboptimal doses. In recent years, model-assisted designs like the Bayesian Optimal Interval (BOIN) have become primary designs for phase I dose-finding studies at the MD Anderson Cancer Center, improving the balance between statistical performance and ease of implementation. To examine long-term trends, this study reviewed 374 completed phase I oncology trials registered on ClinicalTrials.gov between 2014 and 2023. Trials were evaluated based on design methods (3+3, BOIN, mTPI, CRM), dose levels, sample sizes, and other characteristics. A simulation study was conducted to compare the accuracy of MTD selection and patient allocation across designs. The results showed that the 3+3 design remains the most commonly used (61.7%), but BOIN, mTPI, and CRM methods demonstrated higher accuracy in selecting the MTD and treating more patients at the MTD. Notably, BOIN achieved performance comparable to the more complex CRM design while maintaining the simplicity of the 3+3 design, making it a promising alternative for dose-finding in phase I oncology trials.

## 1. Introduction

Phase I clinical trials are essential in the development of anticancer drugs and are the first step in testing new drugs in humans. They focus on assessing a new drug's safety, tolerability, and pharmacokinetics/pharmacodynamics and determining the maximum tolerated dose (MTD) and the recommended dose for Phase II trials. The MTD is often defined as the highest dose patients can receive with tolerable harmful side effects, most of the time referring to the highest dose with a dose-limiting toxicity (DLT) probability closest to the target DLT probability (Yuan et al., 2017).

Phase I clinical trials may have two parts: dose-finding and dose expansion. The goal of dose-finding is to determine the MTD by sequentially testing the toxicity of the study drug from low to high dose levels. The goal of dose expansion is to further confirm the safety of the study drug at MTD determined during the dose-finding and to assess its preliminary efficacy profiles (Yap et al., 2020). Statistical methods/designs are applied to the dose finding process to efficiently explore the dose space and select the MTD.

The statistical methods used for dose-finding fall into three broad categories: rule-based designs, model-based designs, and model-assisted designs (Kurzrock et al., 2021). Strict rule-based designs include 3+3 and Rolling 6. The Continual Reassessment Method (CRM) is a typical example of a model-based design. Model-assisted designs include the Bayesian Optimal Interval (BOIN) Design and Modified Toxicity Probability Interval (mTPI). Here we use 3+3, CRM, mTPI, and BOIN designs as examples to illustrate how they apply to the dose finding process.

### *3+3 Design*

The 3+3 design is a purely rule-based approach in which dose escalation follows pre-specified rules (Dixon & Mood，1946; Storer, 1989). In a 3+3 design, three patients are initially enrolled into the starting dose level to examine how many patients experience a DLT. If none of three patients experience a DLT, then the dose will be escalated to

the next dose level; if one of three patients experience a DLT, then the study will enroll three more patients to the same dose level again; if two or more of three or six patients at that level experience a DLT, that dose level is considered to have exceeded the MTD; further dose escalation will be stopped and the prior dose level will be expanded to six patients. If there is no more than one patient who experiences a DLT among those 6 patients, that dose level is considered the MTD. The MTD is defined as the highest dose level at which six patients were treated and, at most, one patient experienced a DLT.

### *CRM Design*

In 1990, a novel phase I trial design was proposed to improve the efficacy of identifying the MTD - the model-based CRM (O'Quigley et al., 1990). CRM assumes a statistical model for a dose toxicity curve, continuously updating the estimate of the curve based on all current trial data and guiding both dose assignment and MTD selection. Several extensions of CRM have been developed, including dose escalation with overdose control (EWOC) and the Bayesian logistic regression model (BLRM) (Tourneau et al. 2009). Although CRM has better operating characteristics than rule-based designs, implementation in practice is often limited by the need for repeated model fitting (Kurzrock et al. 2021).

### *mTPI Design*

The mTPI and BION designs are model-assisted dose finding methods – a new class of designs that combine the superior performance of model-based methods with the simplicity and transparency of rule-based methods. The mPTI design uses a Bayesian statistical model to calculate the probability that the DLT rate at a given dose falls into one of the three intervals: underdosing, proper (target) dosing, and overdosing. The dose assignment decision will be guided based on posterior probability within each interval and information from previously enrolled patients (Ji and Wang, 2013).

### *BOIN Design*

The BOIN design employs a more direct approach. It compares the observed DLT rate at the current dose with a pair of fixed, prespecified dose escalation and de-escalation boundaries. If the observed DLT rate at a dose level is less than the escalation boundary, the next cohort of patients will be assigned to a higher dose. If the observed DLT rate at a dose level is greater than the de-escalation boundary, the next cohort of patients will be assigned to a lower dose. And if the observed DLT rate at a dose level is between the escalation boundary and the de-escalation boundary, the next cohort of patients will be treated with the same dose level (Liu and Yuan, 2015). In the BOIN design, the escalation and de-escalation boundaries are chosen to minimize decision errors, meaning the risk of incorrectly assigning patients to dose levels that are either too high or too low.

Among all the designs mentioned above, the 3+3 design has traditionally dominated phase I trials due to its simplicity, but it is often criticized for poor operating characteristics, such as low accuracy in identifying the true MTD and inefficient dose escalation. A relatively well-performing model-based design, CRM is sometimes seen as a "black box" due to its statistical complexity, limiting its practical use. In recent years, BOIN has become the primary design used in phase I trials at MD Anderson Cancer Center in Houston, TX, as it provides improved performance (Yuan et al., 2016; Zhou et al., 2018). However, there are no long-term reports on the adoption of dose-finding methods for phase I oncology clinical trials.

The objectives of this study are to learn the basic concepts of phase I clinical trial designs, perform literature reviews of the published phase I clinical trials, summarize the common phase I clinical trial parameters and design methods used in the past 10 years, and conduct simulation studies of novel designs used at MD Anderson to compare which design is the most efficient in finding the MTD.

## 2. Methods and Data Collection

The goal of phase I clinical trials is to determine the safety profiles and select the MTD of the study drug. The primary endpoint of phase I clinical trials is the toxicity. In a phase I clinical trial, we need to define the following parameters: DLT including the toxicity events as well as the assessment time window, target or acceptable toxicity rate and the definition of MTD, patient population, dose levels including the starting dose of the study drug, cohort size in dose-finding, and total sample size.

97   In order to understand the phase I clinical trials being conducted and summarize design parameters, we performed
98   a literature review of the past 10 years completed phase I trials published on ClinicalTrials.gov using the following 8
99   criteria: period of 1/1/2014 - 12/31/2023, disease/condition categorized as cancer, completed study status, phase 1
100  interventional studies, adult patients aged 18 to 64 years, patients of all sexes, results available, and protocol available.
101  Based on the criteria above, we compiled a list of 374 clinical trials and reviewed their protocols to collect data
102  on single or combination drugs, dose-escalation, dose-escalation method (design), dose-expansion, dose levels, sample
103  sizes, toxicity-stopping rules, and phase II processes.
104  Using the parameters collected from the 374 clinical trials, we compared 3+3, BOIN, mTPI, and CRM designs
105  through a simulation study. The simulation was performed using the software and R code developed by the University
106  of Texas MD Anderson Cancer Center (downloaded from https://www.trialdesign.org/ and
107  https://biostatistics.mdanderson.org/SoftwareDownload/). We estimated the percentage of correct MTD selection and
108  the average number of patients allocated to the MTD to evaluate the performance of these designs. Results and findings
109  were analyzed and visualized using Microsoft Excel and statistical software R version 4.2.2 (R Foundation for
110  Statistical Computing).

111  **3.   Results**

112  3.1 Data Summary of 374 Clinical Trials Completed in the Past 10 Years

113  We reviewed 374 Phase I oncology clinical trials completed between January 1, 2014 and December 31, 2023
114  with the criteria mentioned in Methods. Of these 374 clinical trials, 110 trials (29.4%) were single dose studies and
115  50 trials (13.4%) were safety run-in phases. The other 214 trials (57.2%) included the dose-escalation process with
116  three to thirteen dose levels (Figure 1). Of the 214 trials including dose escalation, 132 trials (61.7%) used the 3+3
117  design, 16 trials (7.5%) used CRM, 16 trials (7.5%) used mTPI, only 5 trials (2.3%) used BOIN, and 45 trials (21.0%)
118  used other methods including Bayesian Logistic Regression Model (BLRM), Escalation with Overdose Control
119  (EWOC), Rolling 6, etc. (Figure 2). Of these 214 trials, 74 trials (34.6%) studied a single drug, 133 trials (62.1%)
120  studied a combination of drug treatments, and 7 trials (3.3%) studied both single and combination drugs; 176 trials
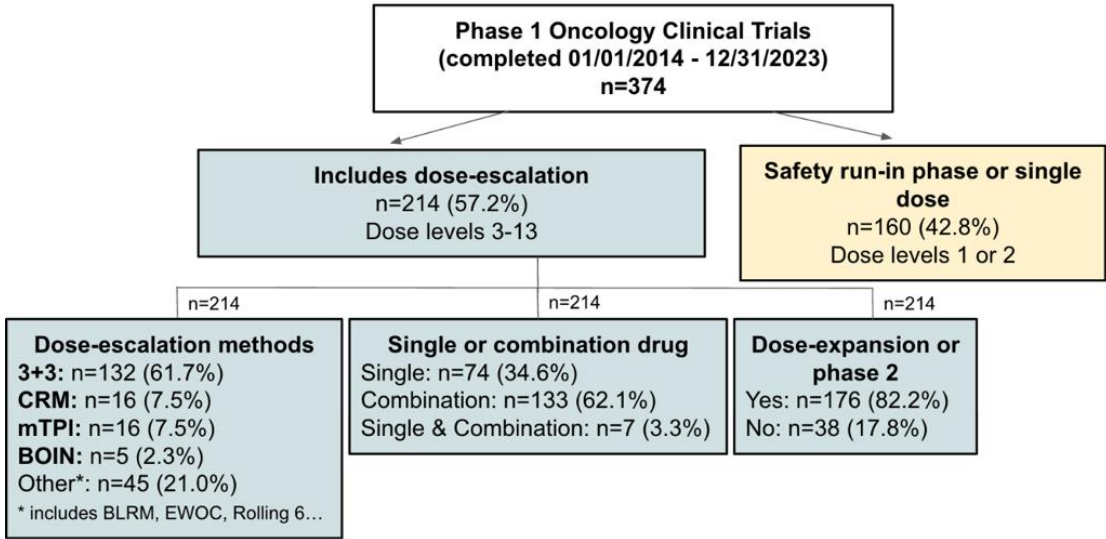121  (82.2%) did conduct dose-expansion or followed with Phase II, while 38 trials (17.8%) did not.
122



123
124
125  Figure 1. Summary of Completed Clinical Trials between 2014 and 2023
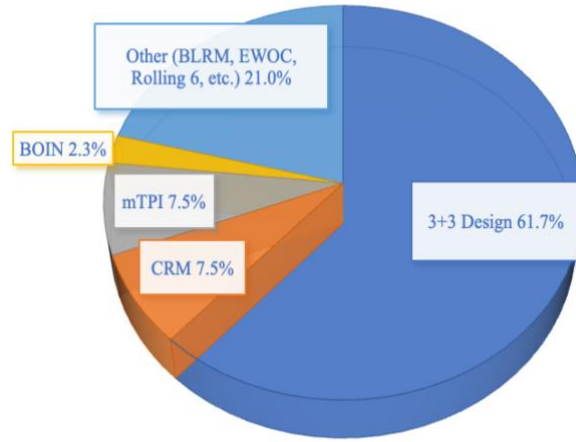126

3

Figure 2. Distribution of Dose-finding Design Methods

Table 1. Sample Size and Dose-Level Summary among 214 Dose-Escalation Trials

| Variable | N | Mean | Median | Std Dev* | Std Error* | Min | Max |
|---|---|---|---|---|---|---|---|
| **Total Sample Size** | 214 | 29.4 | 24.0 | 16.1 | 1.1 | 7 | 106 |
| **Number of Dose Levels** | 214 | 4.8 | 4.0 | 2.1 | 0.1 | 3 | 13 |

*: Std Dev: Standard Deviation; Std Error: Standard Error.

Among the 214 trials with dose-escalation, the average sample size for dose-escalation was about 30 patients, ranging from 7 to 106. The average dose level for dose-escalation was about 5 doses, ranging from 3 to 13 (Table 1).

3.2 Simulation Studies to Evaluate Design Performance

In order to evaluate the properties of the commonly used designs, we conducted simulation studies using an online software and a simulation program developed by the University of Texas MD Anderson Cancer Center. The parameters for the simulation studies were established according to the data collected from the 214 clinical trials. We chose an average sample size of 30 and 5 dose levels as shown in Table 1, using 10 cohorts with a cohort size of 3 patients each. Because most of the studies we reviewed used 0.3 as the target toxicity rate, the simulation target toxicity rate was also set to 0.3 (Le Tourneau et al. 2009). CRM requires a skeleton—the initial set of assumed probabilities of DLT rates for each dose level. The default CRM skeleton provided by the trialdesign.org software was used for the CRM simulation, i.e., we assume that the toxicity rate is 0.20, 0.30, 0.40, 0.50, and 0.59 for Dose 1, 2, 3, 4, and 5, respectively.

We examined 5 representative toxicity scenarios, each defined by the true toxicity rates of the five investigational doses with variation in the location of the MTD (Table 2). In Scenario 1, the first dose (Dose 1) is the MTD at the 0.3 target toxicity rate. In Scenario 2, the MTD is at the second dose. In Scenario 5, the last dose is the MTD. Across Scenarios 1 through 5, the MTD progressively shifts from the lowest to the highest dose level.

Table 2. Toxicity Rates under Simulation Scenario

| | Dose 1 | Dose 2 | Dose 3 | Dose 4 | Dose 5 |
|---|---|---|---|---|---|
| **Scenario 1** | **0.3** | 0.46 | 0.5 | 0.54 | 0.58 |
| **Scenario 2** | 0.16 | **0.3** | 0.47 | 0.54 | 0.6 |
| **Scenario 3** | 0.04 | 0.15 | **0.3** | 0.48 | 0.68 |
| **Scenario 4** | 0.02 | 0.07 | 0.12 | **0.3** | 0.45 |
| **Scenario 5** | 0.02 | 0.06 | 0.1 | 0.13 | **0.3** |

Table 3. Operating Characteristics for 3+3, BOIN, mTPI, and CRM

| | 3+3 | | | | | | BOIN | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dose 1 | Dose 2 | Dose 3 | Dose 4 | Dose 5 | # Pts * | Dose 1 | Dose 2 | Dose 3 | Dose 4 | Dose 5 | # Pts * |
| **Scenario 1** | | | | | | | | | | | | |
| True DLT rate | **0.3** | 0.46 | 0.5 | 0.54 | 0.58 | | **0.3** | 0.46 | 0.5 | 0.54 | 0.58 | |
| Selection % | **35.5** | 6.5 | 1 | 0.1 | 0 | | **69.3** | 15.2 | 2.1 | 0.4 | 0 | |
| # Pts treated | **12.0** | 3.3 | 0.6 | 0.1 | 0.0 | 16.0 | **12.5** | 5.6 | 1.1 | 0.2 | 0.0 | 19.4 |
| **Scenario 2** | | | | | | | | | | | | |
| True DLT rate | 0.16 | **0.3** | 0.47 | 0.54 | 0.6 | | 0.16 | **0.3** | 0.47 | 0.54 | 0.6 | |
| Selection % | 43.2 | **28.1** | 4.9 | 0.6 | 0 | | 29 | **55.1** | 13.1 | 1.5 | 0.1 | |
| # Pts treated | 13.0 | **8.5** | 2.4 | 0.4 | 0.0 | 24.3 | 10.2 | **11.4** | 4.8 | 0.8 | 0.1 | 27.3 |
| **Scenario 3** | | | | | | | | | | | | |
| True DLT rate | 0.04 | 0.15 | **0.3** | 0.48 | 0.68 | | 0.04 | 0.15 | **0.3** | 0.48 | 0.68 | |
| Selection % | 20.7 | 43.6 | **29.2** | 4.8 | 0.1 | | 1.4 | 26.7 | **56.3** | 15.1 | 0.4 | |
| # Pts treated | 7.7 | 11.6 | **7.8** | 2.2 | 0.3 | 29.6 | 5.0 | 9.3 | **10.5** | 4.4 | 0.6 | 29.8 |
| **Scenario 4** | | | | | | | | | | | | |
| True DLT rate | 0.02 | 0.07 | 0.12 | **0.3** | 0.45 | | 0.02 | 0.07 | 0.12 | **0.3** | 0.45 | |
| Selection % | 5 | 13.6 | 45.6 | **29.0** | 6.2 | | 0.2 | 1.8 | 22.2 | **55.7** | 20 | |
| # Pts treated | 4.3 | 5.9 | 10.2 | **7.0** | 2.5 | 29.9 | 3.7 | 4.7 | 8.0 | **9.4** | 4.2 | 30.0 |
| **Scenario 5** | | | | | | | | | | | | |
| True DLT rate | 0.02 | 0.06 | 0.1 | 0.13 | **0.3** | | 0.02 | 0.06 | 0.1 | 0.13 | **0.3** | |
| Selection % | 3.5 | 9.7 | 14.1 | 41.6 | **30.6** | | 0 | 1.2 | 4.4 | 24 | **70.4** | |
| # Pts treated | 3.9 | 5.2 | 5.6 | 8.1 | **7.1** | 29.9 | 3.6 | 4.4 | 5.1 | 7.5 | **9.4** | 29.9 |

| | mTPI | | | | | | CRM | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dose 1 | Dose 2 | Dose 3 | Dose 4 | Dose 5 | # Pts * | Dose 1 | Dose 2 | Dose 3 | Dose 4 | Dose 5 | # Pts * |
| **Scenario 1** | | | | | | | | | | | | |
| True DLT rate | **0.30** | 0.46 | 0.50 | 0.54 | 0.58 | | **0.3** | 0.46 | 0.5 | 0.54 | 0.58 | |
| Selection % | **35.6** | 12.2 | 2.3 | 0.3 | 0.0 | | **55.0** | 17 | 2 | 0 | 0 | |
| # Pts treated | **13.1** | 5.3 | 1.0 | 0.1 | 0.0 | 19.4 | **17.6** | 5.8 | 1.8 | 0.4 | 0.1 | 25.6 |
| **Scenario 2** | | | | | | | | | | | | |
| True DLT rate | 0.16 | **0.30** | 0.47 | 0.54 | 0.60 | | 0.16 | **0.3** | 0.47 | 0.54 | 0.6 | |
| Selection % | 27.8 | **47.5** | 11.7 | 1.7 | 0.0 | | 19 | **58.0** | 18 | 2 | 0 | |
| # Pts treated | 10.5 | **12.1** | 4.0 | 0.6 | 0.1 | 27.3 | 9.8 | **12.2** | 6 | 1.3 | 0.2 | 29.6 |
| **Scenario 3** | | | | | | | | | | | | |
| True DLT rate | 0.04 | 0.15 | **0.30** | 0.48 | 0.68 | | 0.04 | 0.15 | **0.3** | 0.48 | 0.68 | |
| Selection % | 6.7 | 29.8 | **49.3** | 13.2 | 0.4 | | 0 | 17 | **62.0** | 2 | 0 | |
| # Pts treated | 5.0 | 9.8 | **10.9** | 3.7 | 0.4 | 29.8 | 3.8 | 6.8 | **12.3** | 6.2 | 0.9 | 30.0 |
| **Scenario 4** | | | | | | | | | | | | |
| True DLT rate | 0.02 | 0.07 | 0.12 | **0.30** | 0.45 | | 0.02 | 0.07 | 0.12 | **0.3** | 0.45 | |
| Selection % | 1.5 | 4.3 | 28.6 | **48.4** | 17.1 | | 0 | 0 | 13 | **59.0** | 28 | |
| # Pts treated | 3.6 | 4.7 | 8.3 | **9.5** | 4.0 | 30.0 | 3.2 | 3.3 | 5.8 | **10.8** | 6.8 | 30.0 |
| **Scenario 5** | | | | | | | | | | | | |
| True DLT rate | 0.02 | 0.06 | 0.10 | 0.13 | **0.30** | | 0.02 | 0.06 | 0.1 | 0.13 | **0.3** | |
| Selection % | 1.0 | 3.0 | 6.0 | 30.4 | **59.5** | | 0 | 0 | 1 | 17 | **82.0** | |
| # Pts treated | 3.4 | 4.3 | 5.0 | 7.1 | **10.1** | 30.0 | 3.2 | 3.2 | 3.8 | 6.1 | **13.7** | 30.0 |

\* # Pts: Number of patients

Table 3 shows the operating characteristics of 3+3, BOIN, mTPI, and CRM based on 10,000 simulated trials. The selection percentage is the probability of selecting the particular dose level as the MTD. The number of patients treated

159  is the average number of patients treated at the particular dose level. Because 0.3 is the target toxicity rate, the greater
160  the selection percentage and the more patients treated at the target DLT rate, the better the performance of the design.
161  The percentage of correct selection of the MTD and the average number of patients allocated to the MTD were
162  highlighted in Table 3 and visualized in Figure 3. The results in Table 3 indicate that the 3+3 design allocates
163  significantly fewer patients to doses with a 0.3 DLT rate (the true MTD), assigns more patients to lower dose levels,
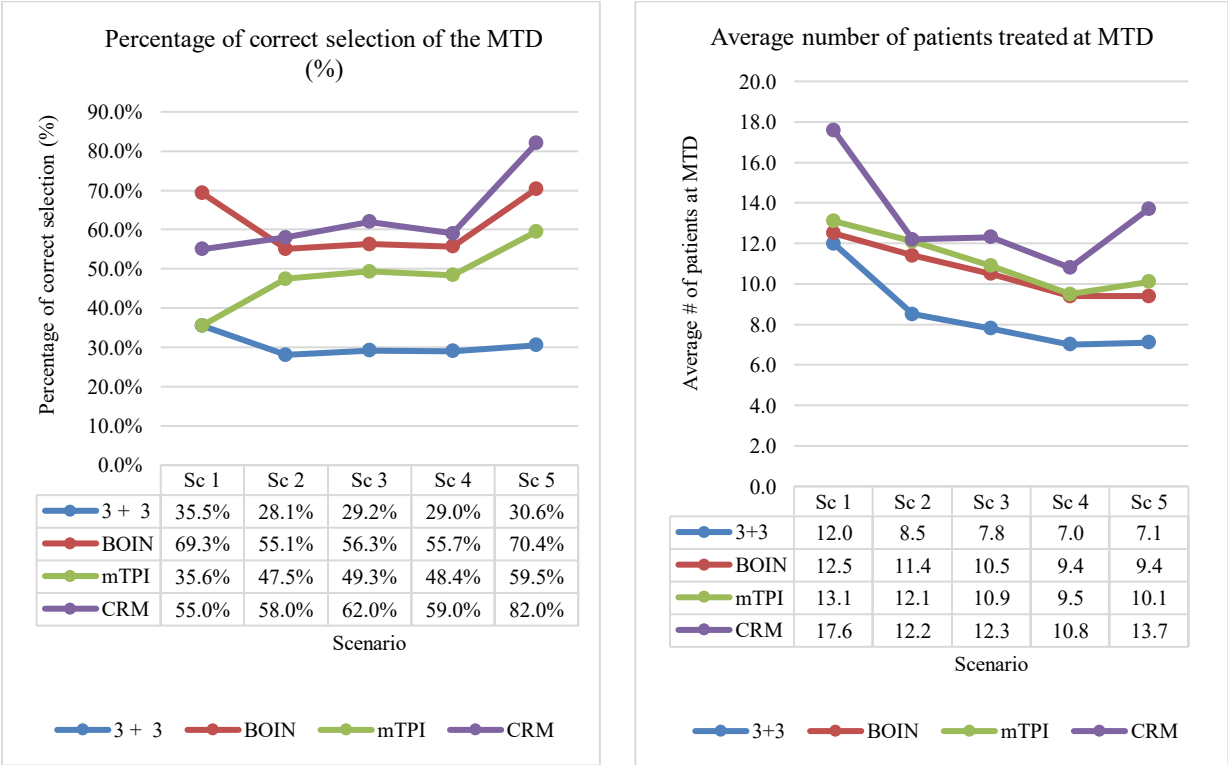164  and rarely explores doses above the target DLT rate.
165
166



### Percentage of correct selection of the MTD (%)

| | Sc 1 | Sc 2 | Sc 3 | Sc 4 | Sc 5 |
|---|---|---|---|---|---|
| 3 + 3 | 35.5% | 28.1% | 29.2% | 29.0% | 30.6% |
| BOIN | 69.3% | 55.1% | 56.3% | 55.7% | 70.4% |
| mTPI | 35.6% | 47.5% | 49.3% | 48.4% | 59.5% |
| CRM | 55.0% | 58.0% | 62.0% | 59.0% | 82.0% |

### Average number of patients treated at MTD

| | Sc 1 | Sc 2 | Sc 3 | Sc 4 | Sc 5 |
|---|---|---|---|---|---|
| 3+3 | 12.0 | 8.5 | 7.8 | 7.0 | 7.1 |
| BOIN | 12.5 | 11.4 | 10.5 | 9.4 | 9.4 |
| mTPI | 13.1 | 12.1 | 10.9 | 9.5 | 10.1 |
| CRM | 17.6 | 12.2 | 12.3 | 10.8 | 13.7 |

167
168  Figure 3. Line Graph Comparing 3+3, BOIN, mTPI, and CRM Performances
169

170  <u>Percentage of correct selection of the MTD (%)</u>
171      As seen with a 0.3 true DLT rate, 3+3 has a relatively low percentage of correctly selecting the MTD in
172  comparison to BOIN, mTPI, and CRM. The low selection percentages reflect the conservative nature of 3+3 and its
173  tendency to select lower doses than the MTD. BOIN has shown performance comparable to CRM and improved
174  accuracy compared to mTPI in selecting the correct MTD.
175
176  <u>Average number of patients allocated to the MTD</u>
177      The performance of the 3+3 design depends on the position of the MTD within the dose range. When the MTD
178  is located at a lower dose level (as seen in scenario 1), the 3+3 design performs similarly to BOIN and mTPI. However,
179  when the MTD is located at higher doses (seen in scenarios 2 -5), the 3+3 design proves to be less efficient than BOIN,
180  mTPI, and CRM, as more patients are assigned to lower suboptimal doses. CRM demonstrates superior performance
181  in allocating more patients to the MTD. BOIN and mTPI show similar performance in patient allocation to the MTD.

182  ## 4.  Discussion

183      Though straightforward and widely used, the 3+3 design is often reported for its poor performance in terms of
184  low selection probability of the correct MTD and more patients being assigned to suboptimal doses.  Despite the
185  development of many clinical trial designs that prove to be more efficient and accurate over the past two decades, the
186  primary reason for the consistent popularity of the 3+3 design as the dominant method used in phase 1 trials is its

simplicity and convenient implementation by clinicians. The dose-escalation and de-escalation rules are fully pre-specified and require no statistical computation, making the design easy to understand and apply in practice. More recently, model-assisted designs such as the BOIN and mTPI designs have been introduced, offering improved performance while retaining the simplicity of pre-specified decision rules. Like the 3+3 design, these methods use predefined escalation and de-escalation boundaries; however, these boundaries are data-driven and depend on the target toxicity rate and cohort size, providing greater flexibility and statistical rigor. Both the literature and our simulation studies demonstrate that model-assisted designs (e.g., BOIN and mTPI) and model-based designs (e.g., CRM) achieve higher accuracy in identifying the true MTD and allocate more patients to the correct dose level, thereby improving both safety and trial efficiency.

From our simulations, BOIN design proved superior to both 3+3 and mTPI in selecting the correct MTD. As a novel, interval-based, and model-assisted approach, BOIN allows for more flexibility in dose adjustments since it is built upon a Bayesian probability model applicable to any cohort size (e.g. 1, 2, 3, 4, or 5 patients per cohort) and any target toxicity rate (e.g. 15%, 20%, 25%, or 30%). In contrast, the 3+3 design uses a rule-based method that follows a rigid algorithm for dose escalation and de-escalation, making the design less adaptable to cohort sizes other than 3 patients per cohort and toxicity rates other than 30%. To enhance patient safety, the BOIN design also includes a dose elimination rule to stop dose escalation if the probability of excessive toxicity is deemed too high. Overall, BOIN is more flexible in specifying target toxicity rates and cohort sizes and demonstrates better performance in selecting the correct MTD and allocating more patients to the MTD dose level than the traditional 3+3 design. Moreover, BOIN's dose escalation and de-escalation decision rules are pre-specified prior to trial conduct, enabling accessible implementation by physicians. Because of its strong performance and ease of use, the BOIN design has been well accepted and has become the most commonly used dose-finding design at MD Anderson Cancer Center and other institutions nationally and internationally.

mTPI is also considered to be more accurate and efficient than the 3+3 design. Both mTPI and BOIN are model-assisted designs and incorporate Bayesian statistical methods to guide dose escalation and de-escalation. Despite their similarities, BOIN employs a simpler and more intuitive approach. It compares the observed DLT rate at the current dose with two fixed, prespecified dose escalation and de-escalation boundaries. These boundaries are calibrated based on a pre-defined target toxicity rate to minimize the risk of incorrect dosing decisions (Liu and Yuan, 2015). This makes BOIN easier to apply in practice and decreases the risk of overdosing compared to mTPI (Chiuzan et al., 2024). Our simulation studies have shown that BOIN treats a similar average number of patients at the MTD as mTPI, but is advantageous in its increased likelihood of determining the MTD correctly compared to mTPI.

Our simulation showed CRM demonstrating superior operating characteristics, including a higher proportion of patients treated at the true MTD and a greater probability of correctly selecting the MTD. Even when using a simple skeleton that assumes each dose level has an equal chance of being the MTD, CRM maintained robust performance across all five simulation scenarios. Despite its statistical advantages, CRM has not replaced the 3+3 design in practice over the past decades. Its adoption has been limited by practical challenges, including the need for repeated model fitting, computational complexity, and a perceived lack of transparency in the dose-escalation process. These barriers have made CRM less accessible to clinicians, especially in settings without dedicated statistical support.

Our simulation results demonstrated that the BOIN, mTPI, and CRM designs had improved accuracy in selecting the MTD compared to the 3+3 design. Due to limited resources, Rolling 6, EWOC, and BLRM were not included in the simulations; however, evaluating these designs remains an important direction for future work. According to other studies (e.g. Zhou et al. 2018), EWOC and BLRM appear to be excessively conservative, ensuring their safety, but they have relatively poor accuracy in finding the MTD. Both mTPI and BOIN offer a balance of efficiency, safety, and ease of implementation, making it easy to use while still providing improvements in accuracy and efficiency. Even though CRM shows improved statistical performance in selecting the correct MTD than other designs, it is inconvenient and difficult to implement by physicians. In contrast, the BOIN design is more intuitive, transparent, and user-friendly, with accessible tools that facilitate its application in practice.

One of the key strengths of this research is the review of 374 clinical trial protocols that spanned over the course of 10 years. This provides sufficient and robust evidence for evaluating and comparing dose-escalation and early safety assessments in clinical trials. This review identifies the gaps between methodological recommendations and actual clinical practice. Another major strength in this research is the use of real-world data collected from the completed clinical trials to establish simulation parameters that align closely with real-world clinical trial characteristics. Additionally, running simulations is a very effective approach to standardize comparisons across methods under identical conditions, which is hard to achieve in real-world settings. This research opportunity deepened my understanding of how statistical methods, simulation programming, and descriptive analysis can be applied to solve real-world problems and assess the effectiveness of dose-finding approaches in oncology clinical

242 trials and health care. An obstacle faced was the lack of transparency in clinical trial protocols. Many studies registered
243 on ClinicalTrials.gov lacked detailed descriptions of their statistical methods and designs, making data collection more
244 challenging.
245     Further research could explore the evolution of trial designs and whether more recent clinical trials are phasing
246 out 3+3 and using more efficient model-assisted methods such as BOIN. To further evaluate and holistically review
247 the performance of different methods, a broader range of parameters could be tested, such as sample size, cohort size,
248 target toxicity rate, and dose levels, better reflecting the diversity of real-world clinical practices.

249 **5. Conclusion**

250     This study investigated the distribution of dose-finding methods in Phase I oncology clinical trials and evaluated
251 the accuracy of dose-escalation designs. After reviewing 374 completed Phase I oncology trials conducted from
252 1/1/2014 - 12/31/2023, we found that the 3+3 design remains the dominant method for dose-escalation, with it being
253 used in approximately 60% of dose-escalation trials. In contrast, more statistically rigorous methods such as BOIN,
254 mTPI, and CRM were used less frequently, despite demonstrating superior performance to 3+3. The simulation results
255 confirmed that these methods, particularly BOIN, are more efficient and accurate than the traditional 3+3 design in
256 selecting the correct MTD and in allocating more patients to optimal dose levels. While our review focused on
257 completed trials, a supplementary search on ClinicalTrials.gov using "BOIN design" as the keyword identified 139
258 registered trials, indicating the growing adoption of the BOIN design in recent clinical practice.

265 **References**

267 Yuan, Y., Nguyen, H.Q., & Thall P.F. (2017). *Bayesian Designs for Phase I-II Clinical Trials.* Chapman and
268 Hall/CRC

270 Yap, T.A., Rodan, J., & Hong D. (2020). *Phase I Oncology Drug Development.* Springer

271 Kurzrock, R., et al. (2021). Moving Beyond 3+3: The Future of Clinical Trial Design. *American Society of Clinical*
272 *Oncology educational book. American Society of Clinical Oncology. Annual Meeting*, 41, e133–e144.
273 https://doi.org/10.1200/EDBK_319783

274 Dixon, W.J., & Mood, A.M. (1946). The statistical sign test. *J Am Stat Assoc. 41:557-566.*

276 Storer, B. E. (1989). Design and analysis of phase I clinical trials. *Biometrics*, *45*(3), 925–937.

278 O'Quigley, J., Pepe, M., & Fisher, L. (1990). Continual reassessment method: a practical design for phase 1 clinical
279 trials in cancer. *Biometrics*, 46(1), 33–48.

281 Le Tourneau, C., Lee, J. J., & Siu, L. L. (2009). Dose escalation methods in phase I cancer clinical trials. *Journal of*
282 *the National Cancer Institute*, 101(10), 708–720. https://doi.org/10.1093/jnci/djp079

284 Ji, Y., & Wang, S. J. (2013). Modified toxicity probability interval design: a safer and more reliable method than the
285 3 + 3 design for practical phase I trials. *Journal of clinical oncology: official journal of the American Society of*
286 *Clinical Oncology*, 31(14), 1785–1791. https://doi.org/10.1200/JCO.2012.45.7903

288 Liu, S., & Yuan, Y. (2015). Bayesian optimal interval designs for phase I clinical trials. *Journal of the Royal*
289 *Statistical Society: Series C.* 64, 507-523.

291   Zhou, H., Yuan, Y., & Nie, L. (2018). Accuracy, Safety, and Reliability of Novel Phase I Trial Designs. *Clinical*
292   *cancer research: an official journal of the American Association for Cancer Research*, 24(18), 4357–4364.
293   https://doi.org/10.1158/1078-0432.CCR-18-0168
294

295   Yuan, Y., et al. (2016). Bayesian Optimal Interval Design: A Simple and Well-Performing Design for Phase I
296   Oncology Trials. *Clinical cancer research: an official journal of the American Association for Cancer Research*,
297   22(17), 4291–4301. https://doi.org/10.1158/1078-0432.CCR-16-0592
298

299   MD Anderson Cancer Center. *Integrated Platform for Designing Clinical Trials.* https://trialdesign.org/
300

301   Quantitative Research Computing at MD Anderson Cancer Center. *Software Download kiosk – Toxicity Probability*
302   *Intervals.* https://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware/Index/72
303

304   Chiuzan, C., & Dehbi, H. M. (2024). The 3 + 3 design in dose-finding studies with small sample sizes: Pitfalls and
305   possible remedies. *Clinical trials (London, England)*, *21*(3), 350–357. https://doi.org/10.1177/17407745241240401