Journal of Research High School

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

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, syliu@mdanderson.org

Received August 3, 2025; Revised January 18, 2026; Accepted March 2, 2026

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 conducted a protocol-based review of 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.

Keywords: Phase I, Oncology clinical trials, Dose-finding studies, 3+3 design, BOIN design

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). Rule-based designs follow fixed escalation rules established in advance that solely depend on the observed number of toxicities at each dose level, without requiring a statistical model. Model-based designs use a statistical model to estimate the dose–toxicity and continuously update dose assignments as patient data accumulate. Model-assisted designs combine the simplicity of rule-based approaches

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

with statistical modeling, using pre-specified decision rules that are informed by probability models to guide dose escalation and de-escalation. 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). This study used 3+3, CRM, mTPI, and BOIN designs as examples to illustrate how they apply to the dose finding process.

1.1 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.

1.2 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).

1.3 mTPI Design

The mTPI and BOIN 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 mTPI 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).

1.4 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, escalation and de-escalation boundaries are selected to minimize decision errors, reducing the risk of assigning patients to dose levels that are 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

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

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 limited long-term reports describing the adoption of dose-finding methods for phase I oncology clinical trials. Adoption is often measured by the proportion of a given dose-escalation method among phase I oncology trials. In this study, the proportion is determined through protocol reviews of completed clinical trials registered on ClinicalTrials.gov. This study summarized the use of dose-escalation design methods over the period from 2014 to 2023 rather than assessing year-by-year changes.

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.

This study hypothesizes that the traditional 3+3 design would remain the most frequently used method across the observed phase I oncology trial protocol reviews. This study further hypothesizes that the CRM, mTPI, and BOIN dose-escalation designs would outperform the 3+3 design, specifically with higher accuracy in identifying the true MTD and improved allocation of patients to the MTD in simulation studies.

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 toxicity. In a phase I clinical trial, the following parameters should be defined: the DLT (specifies what toxicity events are considered unacceptable), the target toxicity rate (the acceptable probability of a DLT for defining the MTD), and basic design features such as the number of dose levels, cohort size, and total sample size.

In order to understand the phase I clinical trials being conducted and summarize design parameters, a protocol-based review was performed using the past 10 years of completed phase I trials registered on ClinicalTrials.gov based on the following 8 criteria: period of 1/1/2014 - 12/31/2023, disease/condition categorized as cancer, completed study status, phase I interventional studies, adult patients aged 18 to 64 years, patients of all sexes, results available, and protocol available. These filters were applied to ensure data completeness and consistent extraction of dose-escalation characteristics.

Based on the criteria above, a list of 374 clinical trials was compiled and their protocols were reviewed to collect data on single or combination drugs, dose-escalation, dose-escalation method (design), dose-expansion, dose levels, sample sizes, toxicity-stopping rules, and phase II processes. The 374 clinical trials included both dose-escalation and non-dose-escalation phase I studies. To minimize selection bias, all completed phase I oncology trials meeting the predefined eligibility criteria were included in the review, regardless of the level of detail provided in the protocol. For trials in which the dose-escalation method was not explicitly stated, the dose-escalation algorithm was carefully reviewed and the trial was classified as the most appropriate known design or "Other" rather than excluding them from the analysis.

Using the parameters collected from the 374 clinical trials, this study compared 3+3, BOIN, mTPI, and CRM designs through a simulation study. The simulation was performed using the software and R code developed by the University of Texas MD Anderson Cancer Center (Yuan et al., n.d.; Ji et al., 2012). This study estimated the percentage of correct MTD selection and the average number of patients allocated to the MTD to evaluate the performance of these designs. Results and findings were analyzed and visualized using Microsoft Excel and statistical software R version 4.2.2 (R Foundation for Statistical Computing).

3. Results

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

This study reviewed 374 Phase I oncology clinical trials completed between January 1, 2014 and December 31, 2023 with the criteria mentioned in Methods. Of these 374 clinical trials, 160 were classified as non-dose-escalation

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

studies: 110 trials (29.4%) were single dose studies and 50 trials (13.4%) were safety run-in phases. The other 214 trials (57.2%) were identified as dose-escalation studies with the criteria of testing at least three dose levels (Figure 1). Among the 214 dose-escalation trials, 132 trials (61.7%) used the 3+3 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%) used Other methods including Bayesian Logistic Regression Model (BLRM), Escalation with Overdose Control (EWOC), Rolling 6, etc., or missing methods (Figure 2). Of these 214 trials, 74 trials (34.6%) studied a single drug, 133 trials (62.1%) studied a combination of drug treatments, and 7 trials (3.3%) studied both single and combination drugs; 176 trials (82.2%) did conduct dose-expansion or followed with Phase II, while 38 trials (17.8%) did not.

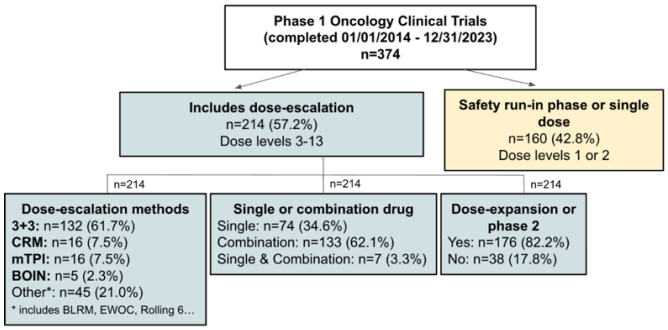


Figure 1. Summary of Completed Clinical Trials between 2014 and 2023.

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).

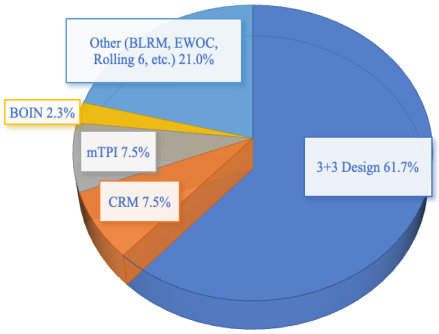


Figure 2. Distribution of Dose-finding Design Methods.

3.2. Simulation Studies to Evaluate Design Performance

In order to evaluate the properties of the commonly used designs, simulation studies were conducted 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. An average sample size of 30 and 5 dose levels were chosen as shown in Table 1, using 10 cohorts with a cohort size of 3 patients each. Because most of the studies 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., the toxicity rate was assumed to be 0.20, 0.30, 0.40, 0.50, and 0.59 for Dose 1, 2, 3, 4, and 5, respectively.

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

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

*: Std Dev: Standard Deviation

This study 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**

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

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

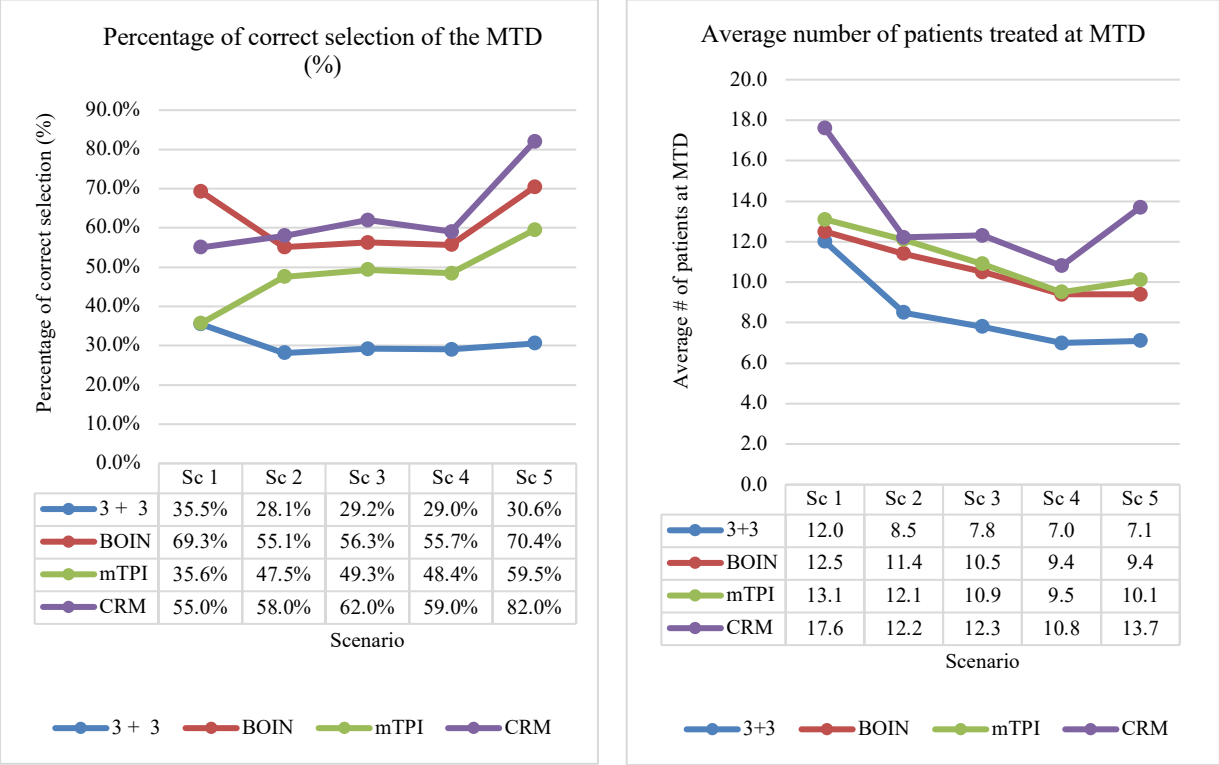
* # Pts: Number of patients

Key takeaways: In Scenarios 1, BOIN performs best in correctly selecting the MTD and CRM performs best in treating most patients at the MTD. In Scenarios 2, 3, 4, and 5, CRM performs best in both correctly selecting the MTD and treating most patients at the MTD. Across all 5 scenarios, 3+3 performed worst in both evaluations.

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

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

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



	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%

	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

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

Percentage of correct selection of the MTD (%)

As seen in Figure 3 (left), with a 0.3 true DLT rate, 3+3 has a relatively low percentage of correctly selecting the MTD in comparison to BOIN, mTPI, and CRM. The low selection percentages reflect the conservative nature of 3+3 and its tendency to select lower doses than the MTD. BOIN has shown performance comparable to CRM and improved accuracy compared to mTPI in selecting the correct MTD.

Average number of patients allocated to the MTD

Figure 3 (right) shows the performance of 3+3 depends on the position of the MTD within the dose range. When the MTD is located at a lower dose level (as seen in scenario 1), the 3+3 design performs similarly to BOIN and mTPI. However, when the MTD is located at higher doses (seen in scenarios 2 -5), the 3+3 design proves to be less efficient than BOIN, mTPI, and CRM, as more patients are assigned to lower suboptimal doses. CRM demonstrates superior performance in allocating more patients to the MTD. BOIN and mTPI show similar performance in patient allocation to the MTD.

4. Discussion

The 3+3 design has remained the most widely used dose-escalation method in Phase I oncology trials, largely due to its simplicity and convenient implementation by clinicians. The dose-escalation and de-escalation rules are fully pre-specified and require no statistical computation, making 3+3 easy to understand and apply in practice. These

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

features have contributed to its continued popularity over several decades.

Despite these advantages, the 3+3 design is often reported for its poor performance in terms of low accuracy in identifying the correct MTD and its tendency to allocate more patients to suboptimal doses. 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 address these limitations, newer dose-finding approaches have been developed. Model-assisted designs, such as BOIN and mTPI, offer 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. Model-based designs, such as CRM, use explicit dose-toxicity models to guide escalation decisions. Both previous studies (Yuan et al., 2016; Zhou et al., 2018) and simulation results demonstrate that these newer designs 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 the 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%). 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). The 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.

The 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.

The 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

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

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 trials and health care.

An obstacle faced was the lack of transparency in clinical trial protocols. Many studies registered on ClinicalTrials.gov lacked detailed descriptions of their statistical methods and designs, making data collection more challenging. The inclusion criteria established to filter clinical trials for review may introduce several limitations. For example, restricting clinical trial samples to adult trials (ages 18-64) likely excluded a substantial number of phase I trials involving pediatric and elderly patients, which may limit the generalizability of the findings. Only reviewing trials with available protocols and results, excluding trials with incomplete or unavailable documentation, may introduce selection bias. As a result, the reviewed clinical trials are representative of trials that meet the inclusion criteria (such as adult trials, available protocols and results, etc.) rather than all phase I oncology trials.

Further research could explore the evolution of trial designs and whether more recent clinical trials are phasing out 3+3 and using more efficient model-assisted methods such as BOIN. To further evaluate and holistically review the performance of different methods, a broader range of parameters could be tested, such as sample size, cohort size, target toxicity rate, and dose levels, better reflecting the diversity of real-world clinical practices.

5. Conclusion

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

Acknowledgment

The author would like to thank Dr. Suyu Liu for her invaluable guidance, support, and mentorship. Her insightful discussions and thoughtful feedback were fundamental in this research journey.

References

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

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

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

Vol. 2026 (5) 413 – 421
ISSN 2688-3651 [online]

Ji, Y., & Herrick, R. (2012, October 5). Software Download kiosk – Toxicity Probability Intervals. Quantitative Research Computing at MD Anderson Cancer Center. https://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware/Index/72

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

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

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

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

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

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

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

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

Yuan, Y., et al. (n.d.). Integrated Platform for Designing Clinical Trials. Trial Design. https://trialdesign.org/

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